FORM PTO-1390 (REV 11-2000) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NUMBER: DIVER1150/WO1
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. APPLICATION NO. (If known, see 37 CFR 1.5) **09/914543**
INTERNATIONAL APPLICATION NO. PCT/US97/08793	INTERNATIONAL FILING DATE 22 May 1997	PRIORITY DATE CLAIMED 22 May 1996
TITLE OF INVENTION ENDOGLUCANASES		
APPLICANT(S) FOR DO/EO/US David W. Lam & Eric J. Mathur		

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 USC 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing 35 USC 371.
3. ☒ This is an express request to begin national examination procedures (35 USC 371(f)). The submission must include items (5), (6), (9), and (21) indicated below.
4. ☐ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☒ is attached hereto (required only if not communicated by the International Bureau).
   b. ☐ has been communicated by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)
6. ☐ An English language translation of the International Application as filed (35 USC 371(c)(2))
   a. ☐ is attached hereto
   b. ☐ has been previously submitted under 35 USC 154(d)(4)
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 USC 371(c)(3)).
9. ☒ An unsigned oath or declaration of the inventor(s) (35 USC 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 USC 371(c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
14. ☐ A SECOND or SUBSEQUENT preliminary amendment.
15. ☐ A substitute specification.
16. ☐ A change of power of attorney and/or address letter.
17. ☐ A computer-readable form of sequence listing in accordance with PCT Rule 13 and 35 USC 1821-1825
18. ☐ A second copy of the published International Application under 35 USC 154(d)(4).
19. ☐ A second copy of the English language translation of the International Application under 35 USC 154(d)(4).
20. ☒ Other items or information: Express Mail Certification; copies of Notification of Receipt of Demand, Notification Concerning Submission of Priority Documents; Petition Under 37 CFR 1.37(b) For Revival Of An Unintentionally Abandoned Application; Statement Under 37 CFR 1.137(b).

U.S. APPLICATION NO (If known, see 37 CFR 1.5) 09/914543	INTERNATIONAL APPLICATION NO	ATTORNEY'S DOCKET NO. DIVER1150WO1

21. __ The following fees are submitted:

		CALCULATIONS	PTO USE ONLY
BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5)): Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2) paid to USPTO and International Search Report not prepared by the EPO or JPO..	$1,000.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by EPO or JPO	$ 860.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO	$ 710.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4).	$ 690.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4).	$ 100.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =		$ 860.00	
Surcharge of $130.00 for furnishing the oath or declaration later than 20__ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).		$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	
Total Claims	16 - 20 =		x $ 18.00	$
Independent Claims	3 - 3 =		x $ 80.00	$
Multiple dependent claim(s) (if applicable)			+ $270.00	$
TOTAL OF ABOVE CALCULATIONS =				$
☒ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$ 430.00
SUBTOTAL =				$ 430.00
Processing fee of $130. for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)). +				$
TOTAL NATIONAL FEE =				$ 430.00
Petition Under 37 CFR 1.37(b) For Revival Of An Unintentionally Abandoned Application Fee				$ 620.00
TOTAL FEES ENCLOSED =				$ 1,050.00
			Amount to be refunded:	$
			charged:	$

a. ☒ A check in the amount of $ 1,050.00 to cover the above fees is enclosed.
b. ☐ Please charge my Deposit Account No._____ in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-1355 . A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

Lisa A. Haile, Ph.D., Reg. No. 38,347

DATE 4/18/01

SEND ALL CORRESPONDENCE TO:
GRAY CARY WARE & FREIDENRICH
4365 Executive Drive, Suite 1600
San Diego, CA 92121-2189

PHONE: 858-677-1456 - FAX: 858-677-1465

ENDOGLUCANASES

Field of the Invention

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polypeptides of the present invention have been identified as endoglucanases and in particular, enzymes having carboxymethyl cellulase activity.

Background

Cellulose, a fibrous, tough, water-insoluble substance is found in the cell walls of plants, particularly, in stalks, stems, trunks and all the woody portions of plant tissues. Cellulose constitutes much of the mass of wood, and cotton is almost pure cellulose. Because cellulose is a linear, unbranched homopolysaccharide of 10,000 to 15,000 D-glucose units, it resembles amylose and the main chains of glycogen. But there is a very important difference, in cellulose, the glucose residues have the beta configuration, whereas in amylose, amylopectin and glycogen, the glucose is in the alpha configuration. The glucose residues in cellulose are linked by (beta 1→4) glycosidic bonds. This difference gives cellulose and amylose very different 3-dimensional structures and physical properties.

Cellulose cannot be used by most animals as a source of stored fuel, because the (beta 1→4) linkages of cellulose are not hydrolyzed by alpha-amylases. Termites readily digest cellulose but only because their intestinal tract harbors a symbiotic microorganism, *trichonympha*, which secretes cellulase, an enzyme that hydrolyses (beta 1→4) linkages between glucose units. The only vertebrates able to use cellulose as food are cattle and other ruminant animals (sheep, goats, camels and giraffes). The extra stomachs "rumens" of these animals teem with bacteria and protists that secrete cellulase.

The enzymatic hydrolysis of cellulose is considered to require the action of both endoglucanases (1,4-beta-D-glucan glucanohydrolase) and exoglucanases (1,4-beta-D-glucan cellobiohydrolase). A synergistic interaction of these enzymes is necessary for the complete hydrolysis of crystalline cellulose. (Caughlin, M.P., Genet. Eng. Rev., 3:39-109 (1985)). For the complete degradation of cellulose (cellulose to glucose), β-glucosidase might be required if the "exo" enzyme does not release glucose. 1,4-β-D-glucan glucohydrolase is another type of "exo" cellulase.

Thermophilic bacteria have received considerable attention as sources of highly active and thermostable cellulolytic and xylanolytic enzymes (Bronneomeier, K. and Staudenbauer, W.L., D.R. Woods (Ed.), The Clostridia and Biotechnology, Butterworth Publishers, Stoneham, MA (1993). Recently, the most extremely thermophilic organotrophic eubacteria presently known have been isolated and characterized. These bacteria, which belong to the genus *thermotoga*, are fermentative microorganisms metabolizing a variety of carbohydrates (Huber, R. and Stetter, K.O., in Ballows, et al., (Ed.), The Procaryotes, 2nd Ed., Springer-Verlaz, New York, pgs. 3809-3819 (1992)).

In Huber *et al.*, 1986, Arch. Microbiol. 144:324-333, the isolation of the bacterium *Thermotoga maritima* is described. *T. maritima* is a eubacterium that is strictly anaerobic, rod- shaped, fermentative, hyperthermophilic, and grows between 55°C and 90°C, with an optimum growth temperature of about 80°C. This eubacterium has been isolated from geothermally heated sea floors in Italy and the Azores. *T. maritima* cells have a sheath-like structure and monotrichous flagellation. *T. maritima* is classified in the eubacterium kingdom by virtue of having murein and fatty acid-containing lipids, diphtheria-toxin-resistant elongation factor 2, an RNA polymerase subunit pattern, and sensitivity to antibiotics.

Since, to date, most organisms identified from the archaeal domain are thermophiles or hyperthermophiles, archaeal bacteria are also considered a fertile source of thermophilic enzymes.

Summary of the Invention

The present invention provides polynucleotides and polypeptides encoded thereby which have been identified as endoglucanase enzymes having carboxymethyl cellulase activity (CMC).

In accordance with one aspect of the present invention, there is provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding enzymes of the present invention including mRNAs, DNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with another aspect of the present invention there are provided isolated nucleic acid molecules encoding mature polypeptides expressed by the DNA contained in ATCC Deposit No. 97516.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptide by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence encoding an enzyme of the present invention, under conditions promoting expression of said enzyme and subsequent recovery of said enzyme.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotide encoding such enzymes for degradation of cellulose for the conversion of plant biomass into fuels and chemicals, for use in detergents, the textile industry, in animal feed, in waste treatment, and in the fruit juice/brewing industry for the clarification and extraction of juices.

In accordance with yet a further aspect of the present invention, there is also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1A-1X show the nucleotide and deduced amino acid sequences the enzymes of the present invention. Sequencing was performed using a 378 automated DNA sequencer (Applied Biosystems, Inc.).

DETAILED DESCRIPTION OF THE INVENTION

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences are ultimately processed to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences. A "promotor sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3′ direction) coding sequence. The promoter is part of the DNA sequence. This sequence region has a start codon at its 3′ terminus. The promoter sequence does include the minimum number of bases where elements necessary to initiate transcription at levels detectable above background. However, after the RNA polymerase binds the sequence and transcription is initiated at the start codon (3′ terminus with a promoter), transcription proceeds downstream in the 3′ direction. Within the promotor sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1) as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

The present invention provides purified thermostable enzymes that catalyze the hydrolysis of the beta 1,4 glycosidic bonds in cellulose to thereby degrade cellulose. An exemplary purified enzyme is an endoglucanase derived from an organism referred herein as "AEPII1a" which is a thermophilic archaeal bacteria which has a very high temperature optimum. The organism is strictly anaerobic, rod-shaped and fermentative, and grows between 55 and 90°C (optimally at 85°C). AEPII1a was discovered in a shallow marine hydrothermal area in Vulcano, Italy. The organism has coccoid cells occurring in singlets or pairs. AEPII1a grows optimally at 85°C and pH 6.5 in a marine medium with cellulose as a substrate and nitrogen in gas phase. This exemplary enzyme is shown in Figure 1A, SEQ ID NO:2.

The polynucleotide encoding SEQ ID NO:2 was originally recovered from a genomic gene library derived from AEPII1a as described below. It contains an open reading frame encoding a protein of 553 amino acid residues.

In one embodiment, the endoglucanase enzyme of SEQ ID NO:2 of the present invention has a molecular weight of about 60.9 kilodaltons as measured by SDS-PAGE gel electrophoresis and an inferred molecular weight from the nucleotide sequence of the gene. This purified enzyme may be used to catalyze the enzymatic degradation of cellulose where desired. The endoglucanase enzyme of the present invention has a very high thermostability and has the closest homology to endo-1,4-beta-glucanase from *Xanthomonas campestris* with 50% identity and 71% similarity at the amino acid level.

In accordance with an aspect of the present invention, there are provided isolated nucleic acid molecules (polynucleotides) which encode for the mature enzymes having the deduced amino acid sequence of Figure 1A-X.

This invention, in addition to the isolated nucleic acid molecule encoding an endoglucanase enzyme disclosed in Figure 1 (SEQ ID NO:1), also provides substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are

capable of hybridizing under stringent conditions, hereinafter described, to SEQ ID NO:1; or (ii) they encode DNA sequences which are degenerate to SEQ ID NO:1. Degenerate DNA sequences encode the amino acid sequence of SEQ ID NO:2, but have variations in the nucleotide coding sequences. As used herein, "substantially similar" refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially similar can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially similar can be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating a nucleic acid molecule encoding an endoglucanase enzyme is to probe a genomic gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, Ausubel F.M. *et al.* (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that SEQ ID NO:1, or fragments thereof (comprising at least 15 contiguous nucleotides), is a particularly useful probe. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NO:1 (*i.e.*, comprising at least 15 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acid is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH_2PO_4, pH 7.0, 5.0 mM Na_2EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/mL polyriboadenylic acid. Approximately 2 X 10^7 cpm (specific activity 4-9 X 10^8 cpm/ug) of ^{32}P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na_2EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm-10°C for the oligo-nucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. *See* J. Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2d Ed. 1989) (Cold Spring Harbor Laboratory) which is hereby incorporated by reference in its entirety.

"Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NO:1). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention also relates to polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the changes do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the enzyme encoded by the reference polynucleotide (SEQ ID NO:1). In a preferred aspect of the invention these enzymes retain the same biological action as the enzyme encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other animal sources or to screen such sources for related sequences.

The present invention provides substantially pure endoglucanase enzymes. The term "substantially pure" is used herein to describe a molecule, such as a polypeptide (*e.g.*, an

endoglucanase polypeptide, or a fragment thereof) that is substantially free of other proteins, lipids, carbohydrates, nucleic acids, and other biological materials with which it is naturally associated. For example, a substantially pure molecule, such as a polypeptide, can be at least 60%, by dry weight, the molecule of interest. The purity of the polypeptides can be determined using standard methods including, *e.g.*, polyacrylamide gel electrophoresis (*e.g.*, SDS-PAGE), column chromatography (*e.g.*, high performance liquid chromatography (HPLC)), and amino-terminal amino acid sequence analysis.

Endoglucanase polypeptides included in the invention can have one of the amino acid sequences of Endoglucanases shown in Figures 1A through 1X (SEQ ID NO:NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48), for example, the amino acid sequence of AEPII1a (SEQ ID NO:2). Endoglucanase polypeptides, such as those isolated from AEPII1a , can be characterized by catalyzing the hydrolysis of the beta 1,4 glycosidic bonds in cellulose.

Also included in the invention are polypeptides having sequences that are "substantially identical" to the sequence of an endoglucanase polypeptide, such as one of SEQ ID NO:NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48, *e.g.*, SEQ ID NO:2. A "substantially identical" amino acid sequence is a sequence that differs from a reference sequence only by conservative amino acid substitutions, for example, substitutions of one amino acid for another of the same class (*e.g.*, substitution of one hydrophobic amino acid, such as isoleucine, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid, or glutamine for asparagine), or by one or more non-conservative substitutions, deletions, or insertions, provided that the polypeptide retains at least one endoglucanase-specific activity or an endoglucanase-specific epitope. For example, one or more amino acids can be deleted from an endoglucanase polypeptide, resulting in modification of the structure of the polypeptide, without significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for

endoglucanase biological activity, can be removed. Such modifications can result in the development of smaller active endoglucanase polypeptides.

Other endoglucanase polypeptides included in the invention are polypeptides having amino acid sequences that are at least 50% identical to the amino acid sequence of an endoglucanase polypeptide, such as any of endoglucanases in SEQ ID NO:NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48, *e.g.*, SEQ ID NO:2. The length of comparison in determining amino acid sequence homology can be, for example, at least 15 amino acids, for example, at least 20, 25, or 35 amino acids. Homology can be measured using standard sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705; also see Ausubel, *et al., supra*).

The invention also includes fragments of endoglucanase polypeptides that retain at least one endoglucanase-specific activity or epitope. Endoglucanase activity can be assayed by examining the catalysis of beta 1,4 glycosidic bonds in cellulose. For example, an endoglucanase polypeptide fragment containing, *e.g.*, at least 8-10 amino acids can be used as an immunogen in the production of endoglucanase-specific antibodies. The fragment can contain, for example, an amino acid sequence that is conserved in endoglucanases, and this amino acid sequence can contain amino acids that are conserved in endoglucanases. Such fragments can easily be identified by comparing the sequences of endoglucanases found in Figures 1A-1X. In addition to their use as peptide immunogens, the above-described endoglucanase fragments can be used in immunoassays, such as ELISAs, to detect the presence of endoglucanase-specific antibodies in samples.

The endoglucanase polypeptides of the invention can be obtained using any of several standard methods. For example, endoglucanase polypeptides can be produced in a standard recombinant expression systems (see below), chemically synthesized (this approach may be

limited to small endoglucanase peptide fragments), or purified from organisms in which they are naturally expressed.

The invention also provides isolated nucleic acid molecules that encode the endoglucanase polypeptides described above, as well as fragments thereof. For example, nucleic acids that encode any of SEQ ID NO:NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48, are included in the invention. These nucleic acids can contain naturally occurring nucleotide sequences, or sequences that differ from those of the naturally occurring nucleic acids that encode endoglucanases, but encode the same amino acids, due to the degeneracy of the genetic code. The nucleic acids of the invention can contain DNA or RNA nucleotides, or combinations or modifications thereof. Exemplary nucleic acids of the invention are shown in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, and 47.

By "isolated nucleic acid" is meant a nucleic acid, *e.g.*, a DNA or RNA molecule, that is not immediately contiguous with the 5' and 3' flanking sequences with which it normally is immediately contiguous when present in the naturally occurring genome of the organism from which it is derived. The term thus describes, for example, a nucleic acid that is incorporated into a vector, such as a plasmid or viral vector; a nucleic acid that is incorporated into the genome of a heterologous cell (or the genome of a homologous cell, but at a site different from that at which it naturally occurs); and a nucleic acid that exists as a separate molecule, *e.g.*, a DNA fragment produced by PCR amplification or restriction enzyme digestion, or an RNA molecule produced by *in vitro* transcription. The term also describes a recombinant nucleic acid that forms part of a hybrid gene encoding additional polypeptide sequences that can be used, for example, in the production of a fusion protein.

The nucleic acid molecules of the invention can be used as templates in standard methods for production of endoglucanase gene products (*e.g.*, endoglucanase RNAs and endoglucanase polypeptides). In addition, the nucleic acid molecules that encode

endoglucanase polypeptides (and fragments thereof) and related nucleic acids, such as (1) nucleic acids containing sequences that are complementary to, or that hybridize to, nucleic acids encoding endoglucanase polypeptides, or fragments thereof (*e.g.*, fragments containing at least 12, 15, 20, or 25 nucleotides); and (2) nucleic acids containing sequences that hybridize to sequences that are complementary to nucleic acids encoding endoglucanase polypeptides, or fragments thereof (*e.g.*, fragments containing at least 12, 15, 20, or 25 nucleotides); can be used in methods focused on their hybridization properties. For example, as is described in further detail below, such nucleic acid molecules can be used in the following methods: PCR methods for synthesizing endoglucanase nucleic acids, methods for detecting the presence of an endoglucanase nucleic acid in a sample, screening methods for identifying nucleic acids encoding new endoglucanase family members.

The invention also includes methods for identifying nucleic acid molecules that encode members of the endoglucanase polypeptide family in addition to SEQ ID NO:NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48. In these methods, a sample, *e.g.*, a nucleic acid library, such as a cDNA library, that contains a nucleic acid encoding an endoglucanase polypeptide is screened with an endoglucanase-specific probe, *e.g.*, an endoglucanase-specific nucleic acid probe. Endoglucanase-specific nucleic acid probes are nucleic acid molecules (*e.g.*, molecules containing DNA or RNA nucleotides, or combinations or modifications thereof) that specifically hybridize to nucleic acids encoding endoglucanase polypeptides, or to complementary sequences thereof. The term "endoglucanase-specific probe," in the context of this method of invention, refers to probes that bind to nucleic acids encoding endoglucanase polypeptides, or to complementary sequences thereof, to a detectably greater extent than to nucleic acids encoding other enzymes, or to complementary sequences thereof.

The invention facilitates production of endoglucanase-specific nucleic acid probes. Methods for obtaining such probes can be designed based on the amino acid sequences shown in Figure 1. The probes, which can contain at least 12, *e.g.*, at least 15, 25, 35, 50, 100, or 150 nucleotides, can be produced using any of several standard methods (see, *e.g.*, Ausubel,

et al., supra). For example, preferably, the probes are generated using PCR amplification methods. In these methods, primers are designed that correspond to endoglucanase-conserved sequences (see Figure 1), which can include endoglucanase-specific amino acids, and the resulting PCR product is used as a probe to screen a nucleic acid library, such as a cDNA library.

In accordance with another aspect of the present invention, there is provided an isolated polynucleotide encoding an exemplary enzyme of the present invention (SEQ ID NO:1) which has been deposited with an appropriate depository for the deposit of biological material. The deposited material is a pQET (Qiagen, Inc.) plasmid comprising the DNA of Figure 1A. The deposit has been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA, on April 22, 1996 and assigned ATCC Deposit No. 97516.

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the deposit of micro-organisms for purposes of patent procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and are not an admission that a deposit be required under 35 U.S.C. §112. The sequences of the polynucleotides contained in the deposited materials, as well as the amino acid sequences of the polypeptides encoded thereby, are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The coding sequences for the endoglucanase enzymes of the present invention were identified by preparing an AEPII1a genomic DNA library, for example, and screening the library for the clones having endoglucanase activity. Such methods for constructing a genomic gene library are well-known in the art. One means, for example, comprises shearing DNA isolated from AEPII1a by physical disruption. A small amount of the sheared DNA is checked on an agarose gel to verify that the majority of the DNA is in the desired size range

(approximately 3-6 kb). The DNA is then blunt ended using Mung Bean Nuclease, incubated at 37°C and phenol/chloroform extracted. The DNA is then methylated using Eco RI Methylase. Eco RI linkers are then ligated to the blunt ends through the use of T4 DNA ligase and incubation at 4°C. The ligation reaction is then terminated and the DNA is cut-back with Eco RI restriction enzyme. The DNA is then size fractionated on a sucrose gradient following procedures known in the art, for example, Maniatis, T., *et al.*, Molecular Cloning, Cold Spring Harbor Press, New York, 1982, which is hereby incorporated by reference in its entirety.

A plate assay is then performed to get an approximate concentration of the DNA. Ligation reactions are then performed and 1 µl of the ligation reaction is packaged to construct a library. Packaging, for example, may occur through the use of purified λgt11 phage arms cut with EcoRI and DNA cut with EcoRI after attaching EcoRI linkers. The DNA and λgt11 arms are ligated with DNA ligase. The ligated DNA is then packaged into infectious phage particles. The packaged phages are used to infect E. coli cultures and the infected cells are spread on agar plates to yield plates carrying thousands of individual phage plaques. The library is then amplified.

In a preferred embodiment, the enzyme of the present invention, was isolated from an AEPII1a library by the following technique:

1. λgt11 AEPII1a library was plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates (~4,800 plaque forming units/plate) in E.coli Y1090 host with LB agarose containing 1mM IPTG as top agarose. The plates were incubated at 37°C overnight.

2. Plates were chilled at 4°C for one hour.

3. The plates were overlayed with Duralon membranes (Stratagene) at room temperature for one hour and the membranes were oriented and lifted off the plates and stored at 4°C.

4. The top agarose layer was removed and plates were incubated at 72°C for ~3 hours.

5. The plate surface was rinsed with NaCl.

6. The plate was stained with 0.1% Congo Red for 15 minutes.

7. The plate was destained with 1M NaCl.

8. The putative positives identified on plate were isolated from the Duralon membrane (positives are identified by clearing zones around clones). The phage was eluted from the membrane by incubating in 500μl SM + 25μl $CHCl_3$ to elute.

9. Insert DNA was subcloned into pBluescript II SK(+) cloning vector (Stratagene), and subclones were reassayed for CMCase activity using the following protocol:

i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.

ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.

iii) Incubate at 72°C for 2 hours.

iv) Stain with 0.1% Congo Red for 15 minutes.

v) Destain with 1M NaCl for 15 minutes.

vi) Identify positives by clearing zone around clone.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological

activity. Probes of this type have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns.

The isolated nucleic acid sequences and other enzymes may then be measured for retention of biological activity characteristic to the enzyme of the present invention, for example, in an assay for detecting enzymatic endoglucanase activity. Such enzymes include truncated forms of endoglucanase, and variants such as deletion and insertion variants.

Examples of such assays include an assay for the detection of endoglucanase activity based on specific interaction of direct dyes such as Congo red with polysaccharides. This colorant reacts with beta-1,4-glucans causing a visible red shift (Wood, P.J., Carbohydr. Res., 85:271 (1980) and Wood, P.J., Carbohydr. Res., 94:c19 (1981)). The preferred substrate for the test is carboxymethylcellulose (CMC) which can be obtained from different sources (Hercules Inc., Wilmington, DE, Type 4M6F or Sigma Chemical Company, St. Louis, MO, Medium Viscosity). The CMC is incorporated as the main carbon sources into a minimal agar medium in quantities of 0.1-1.0%. The microorganisms can be screened directly on these plates, but the replica plating technique from a master plate is preferable since the visualization of the activity requires successive flooding with the reagents, which would render the reisolation of active colonies difficult. Such endoglucanase-producing colonies are detectable after a suitable incubation time (1-3 days depending on the growth), by flooding the plate with 10 ml of a 0.1% aqueous solution of Congo Red. The coloration is terminated after 20 minutes by pouring off the dye and flooding the plates with 10 ml of 5M NaCl solution (commercial salt can be used). After an additional 20 minutes, the salt solution is discarded and endoglucanase activity is revealed by a pale-orange zone around the active microorganisms. In some cases, these zones can be enhanced by treating the plates with 1 N acetic acid, causing the dye to change its color to blue.

The same technique can be used as a cup-plate diffusion assay with excellent sensitivity for the determination of endoglucanase activity in culture filtrates or during enzyme purification steps (Carger, J.H., Anal. Biochem., 153:75 (1986)). See generally, Methods for Measuring Cellulase Activities, Methods in Enzymology, Vol. 160, pgs. 87-116.

The enzyme of the present invention has enzymatic activity with respect to the hydrolysis of the beta 1,4 glycosidic bonds in carboxymethylcellulose, since the halos discussed above are shown around the colonies.

The polynucleotide of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature enzyme may be identical to the coding sequences shown in Figure 1 and/or that of the deposited clone (SEQ ID NO:1), or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzyme as the DNA of Figure 1 (*e.g.*, SEQ ID NO:1).

The polynucleotide which encodes for the mature enzyme of Figure 1 (*e.g.*, SEQ ID NO:2) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzyme having the deduced amino acid sequence of Figure 1 (*e.g.*, SEQ ID NO:2). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzyme as shown in Figure 1 as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzyme of Figure 1. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence shown in Figure 1. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

The present invention also includes polynucleotides, wherein the coding sequence for the mature enzyme may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of an enzyme from a host cell, for example, a leader sequence which functions to control transport of an enzyme from the cell. The enzyme having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the enzyme. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature enzyme, or for an enzyme having a prosequence or for an enzyme having both a prosequence and a presequence (leader sequence).

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figure 1.

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotide of SEQ ID NO:1, for example, for recovery of the polynucleotide or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzyme of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, or 48 as well as fragments thereof, which fragments have at least 30 bases and preferably at least 50 bases and to enzymes encoded by such polynucleotides.

The present invention further relates to a enzyme which has the deduced amino acid sequence of Figure 1, as well as fragments, analogs and derivatives of such enzyme.

The terms "fragment," "derivative" and "analog" when referring to the enzyme of Figure 1 means a enzyme which retains essentially the same biological function or activity as

such enzyme. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzyme of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzyme of Figure 1 may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include an enzyme of Figure 1A-1X (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to an enzyme of Figure 1A-1X and more preferably at least 90% similarity (more preferably at least 90% identity) to an enzyme of Figure 1A-1X and still more preferably at least 95% similarity (still more preferably at least 95% identity) to an enzyme of Figure 1A-1X and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme. Similarity may be determined by procedures which are well-known in the art, for example, a BLAST program (Basic Local Alignment Search Tool at the National Cneter for Biological Information).

A variant, i.e. a "fragment", "analog" or "derivative" enzyme, and reference enzyme may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors containing the polynucleotides of this invention. Such vectors may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction

endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, Streptomyces, *Bacillus subtilis*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, *etc.* The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this

embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBluescript II (Stratagene); pTRC99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene) pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the

present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium*

and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

The enzyme of this invention may be employed for any purpose in which such enzyme activity is necessary or desired. In a preferred embodiment the enzyme is employed for catalyzing the hydrolysis of cellulose. The degradation of cellulose may be used for the conversion of plant biomass into fuels and chemicals.

The enzyme of the present invention may also be employed in the detergent and textile industry, in the production of animal feed, in waste treatment and in the fruit juice/brewing industry for the clarification and extraction of juices.

In a preferred embodiment, the enzyme of the present invention is a thermostable enzyme which is stable to heat and is heat resistant and catalyzes the enzymatic hydrolysis of cellulose, *i.e.*, the enzyme is able to renature and regain activity after a brief (*i.e.*, 5 to 30 seconds), or longer period, for example, minutes or hours, exposure to temperatures of 80°C to 105°C and has a temperature optimum above 60°C.

The enzymes, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in Enzymology*, Vol 160, pp. 87-116, which is

hereby incorporated by reference in its entirety. Antibodies may also be employed as a probe to screen gene libraries generated from this or other organisms to identify this or cross reactive activities.

Isolation and purification of polypeptides produced in the systems described above can be carried out using conventional methods, appropriate for the particular system. For example, preparative chromatography and immunological separations employing antibodies, such as monoclonal or polyclonal antibodies, can be used.

The term "antibody," as used herein, refers to intact immunoglobulin molecules, as well as fragments of immunoglobulin molecules, such as Fab, Fab', $(Fab')_2$, Fv, and SCA fragments, that are capable of binding to an epitope of an endoglucanase polypeptide. These antibody fragments, which retain some ability to selectively bind to the antigen (*e.g.*, an endoglucanase antigen) of the antibody from which they are derived, can be made using well known methods in the art (see, *e.g.*, Harlow and Lane, *supra*), and are described further, as follows.

(1) A Fab fragment consists of a monovalent antigen-binding fragment of an antibody molecule, and can be produced by digestion of a whole antibody molecule with the enzyme papain, to yield a fragment consisting of an intact light chain and a portion of a heavy chain.

(2) A Fab' fragment of an antibody molecule can be obtained by treating a whole antibody molecule with pepsin, followed by reduction, to yield a molecule consisting of an intact light chain and a portion of a heavy chain. Two Fab' fragments are obtained per antibody molecule treated in this manner.

(3) A $(Fab')_2$ fragment of an antibody can be obtained by treating a whole antibody molecule with the enzyme pepsin, without subsequent reduction. A $(Fab')_2$ fragment is a dimer of two Fab' fragments, held together by two disulfide bonds.

(4) An Fv fragment is defined as a genetically engineered fragment containing the variable region of a light chain and the variable region of a heavy chain expressed as two chains.

(5) A single chain antibody ("SCA") is a genetically engineered single chain molecule containing the variable region of a light chain and the variable region of a heavy chain, linked by a suitable, flexible polypeptide linker.

As used in this invention, the term "epitope" refers to an antigenic determinant on an antigen, such as an endoglucanase polypeptide, to which the paratope of an antibody, such as an endoglucanase-specific antibody, binds. Antigenic determinants usually consist of chemically active surface groupings of molecules, such as amino acids or sugar side chains, and can have specific three-dimensional structural characteristics, as well as specific charge characteristics.

As is mentioned above, antigens that can be used in producing endoglucanase-specific antibodies include endoglucanase polypeptides, *e.g.*, any of the endoglucanases shown in Figures 1A-1X polypeptide fragments. The polypeptide or peptide used to immunize an animal can be obtained by standard recombinant, chemical synthetic, or purification methods. As is well known in the art, in order to increase immunogenicity, an antigen can be conjugated to a carrier protein. Commonly used carriers include keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The coupled peptide is then used to immunize the animal (*e.g.*, a mouse, a rat, or a rabbit). In addition to such carriers, well known adjuvants can be administered with the antigen to facilitate induction of a strong immune response.

Endoglucanase-specific polyclonal and monoclonal antibodies can be purified, for example, by binding to, and elution from, a matrix containing an endoglucanase polypeptide, *e.g.*, the endoglucanase polypeptide (or fragment thereof) to which the antibodies were raised. Additional methods for antibody purification and concentration are well known in the art and can be practiced with the endoglucanase-specific antibodies of the invention (see, for example, C-oligan, *et al.*, Unit 9, *Current Protocols in Immunology*, Wiley Interscience, 1994).

Anti-idiotype antibodies corresponding to endoglucanase-specific antigens are also included in the invention, and can be produced using standard methods. These antibodies are raised to endoglucanase-specific antibodies, and thus mimic endoglucanase-specific epitopes.

The members of a pair of molecules (*e.g.*, an antibody-antigen pair or a nucleic acid pair) are said to "specifically bind" to each other if they bind to each other with greater affinity than to other, non-specific molecules. For example, an antibody raised against an antigen to which it binds more efficiently than to a non-specific protein can be described as specifically binding to the antigen. (Similarly, a nucleic acid probe can be described as specifically binding to a nucleic acid target if it forms a specific duplex with the target by base pairing interactions (see above).)

The present invention is further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In one aspect of the invention, a method for producing an endoglucanase enzyme, such as those shown in Figures 1A-1X, is provided. The method includes growing a host cell which contains a polynucleotide encoding the enzyme (*e.g.*, SEQ ID NO: NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, or 47), under conditions which allow the expression of the nucleic acid, and isolating the enzyme encoded by the nucleic acid. Methods of culturing the host cell are described in the Examples and are known by those of skill in the art.

In another embodiment, the invention provides a method for degrading carboxymethylcellulose. The method includes contacting carboxymethylcellulose with a degrading effective amount of an enzyme of the invention, such as the enzyme shown in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, or 48. The term "degrading effective" amount refers to the amount of enzyme which is required to degrade at least 50% of the carboxymethylcellulose, as compared to carboxymethylcellulose

not contacted with the enzyme. Preferably, at least 80% of the carboxymethylcellulose is degraded.

In another embodiment, the invention provides a method for hydrolyzing the beta 1,4 glycosidic bond in cellulose, the method including administering an effective amount of an enzyme of the invention (*e.g.*, SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, or 48) to cellulose, to hydrolyze the glycosidic bond. An "effective" amount refers to the amount of enzyme which is required to hydrolyze at least 50% of the glycosidic bonds, as compared to carboxymethylcellulose not contacted with the enzyme. Preferably, at least 80% of the glycosidic bonds are hydrolyzed.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary

in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is generally performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, Nucleic Acids Res., 8:4057 (1980), for example.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides may or may not have a 5' phosphate. Those that do not will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Sambrook, Fritsch and Maniatus, 1989. The following examples are intended to illustrate, but not to limit, the invention. While the procedures described in the examples are typical of those that can be used to carry out certain aspects of the invention, other procedures known to those skilled in the art can also be used. The following materials and methods were used in carrying out the experiments described in the examples.

Example 1

Bacterial Expression and Purification of Endoglucanase

An AEPII1a genomic library was constructed in the Lambda gt11 cloning vector (Stratagene Cloning Systems). The library was screened in Y1090 E. coli cells (Stratagene) for endoglucanase activity and a positive clone was identified and isolated. DNA of this clone was used as a template in a 100 ul PCT reaction using the following primer sequences:
5' primer: AATAGCGGCCGCAAGCTTATCGACGGTTTCCATATGGGGATTGGTG (SEQ ID NO:49). 3' primer: AATAGCGGCCGCGGATCCAGACCAACTGG TAATGGTAGCGAC (SEQ ID NO:50).

The PCR reaction product was purifed and digested with Not I restriction enzyme. The digested product was subcloned into the pBluescript II SK cloning vector (Stratagene) and sequenced. The sequence information was used in the generation of primer sequences which were subsequently used to PCR amplify the target gene encoding the endoglucanase. The primer sequences used were as follows:
5' primer: TTTATTCAATTGATTAAAGAGGAGAAATTAACTATGATAAACGTTGC AACGGGAGAGGAG (SEQ ID NO:51) and
3' primer: TTTATTGGATCCTACTTTGTGTCAACGAAGTATCC (SEQ ID NO:52).

The amplification product was digested with the restriction enzymes MfeI and BamHI. The digested product was then ligated to pQET cloning vector, a modified form of a pQE vector (Qiagen, Inc.) which was previously digested with BamHI and EcoRI compatible with MfeI. The pQE vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The amplified sequences were inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the *E. coli* strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Positive recombinant

transformants were identified as having thermostable CMCase/endoglucanase activity by the assay described above. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 ($O.D.^{600}$) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described. It is to be understood that, while the invention has been described with reference to the above detailed description, the foregoing description is intended to illustrate, but not to limit, the scope of the invention. Other aspects, advantages, and modifications of the invention are within the scope of the following claims. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

What Is Claimed Is:

1. Substantially pure endoglucanase having an amino acid sequence selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48.

2. An isolated polynucleotide sequence encoding an endoglucanase of claim 1.

3. An isolated polynucleotide selected from the group consisting of:
   a) SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, and 47;
   b) SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, and 47, wherein T can also be U;
   c) nucleic acid sequences complementary to a) and b); and
   d) fragments of a), b), or c) that are at least 15 bases in length and that will selectively hybridize to DNA which encodes the amino acid sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, and 48.

4. The polynucleotide of claim 2, wherein the polynucleotide is isolated from a prokaryote.

5. An expression vector including the polynucleotide of claim 2.

6. The vector of claim 5, wherein the vector is a plasmid.

7. The vector of claim 5, wherein the vector is a virus-derived.

8. A host cell transformed with the vector of claim 5.

9. The host cell of claim 8, wherein the cell is prokaryotic.

10. Antibodies that bind to the polypeptide of claim 1.

11. The antibodies of claim 10, wherein the antibodies are polyclonal.

12. The antibodies of claim 10, wherein the antibodies are monoclonal.

13. An enzyme comprising a member selected from the group consisting of:
    a) an enzyme comprising an amino acid sequence which is at least 70% identical to the amino acid sequence set forth in SEQ ID NO:2; and
    b) an enzyme which comprises at least 30 amino acid residues to the enzyme of a).

14. A method for producing an enzyme comprising growing a host cell of claim 8 under conditions which allow the expression of the nucleic acid and isolating the enzyme encoded by the nucleic acid.

15. A method for degrading carboxymethylcellulose comprising contacting carboxymethylcellulose with a degrading effective amount of the enzyme of claim 1.

16. A method for hydrolyzing the beta 1,4 glycosidic bond in cellulose comprising contacting an effective amount of the enzyme of claim 1 with cellulose to hydrolyze the glycosidic bond.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 16/40, C12N 1/21, 5/10, 9/42, 15/56, 15/63, C12P 21/02, C12S 3/02

A1

(11) International Publication Number: **WO 97/44361**

(43) International Publication Date: 27 November 1997 (27.11.97)

(21) International Application Number: PCT/US97/08793

(22) International Filing Date: 22 May 1997 (22.05.97)

(30) Priority Data:
08/651,572 22 May 1996 (22.05.96) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/651,572 (CIP)
Filed on 22 May 1996 (22.05.96)

(71) Applicant *(for all designated States except US)*: RECOMBINANT BIOCATALYSIS, INC. [US/US]; 505 West Coast Boulevard, La Jolla, CA 92037 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US).

(74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ENDOGLUCANASES

(57) Abstract

The invention provides a purified thermostable enzyme derived from the archael bacterium AEPH1a. The enzyme has a molecular weight of about 60.9 kilodaltons and has cellulase activity (SEQ ID NO:2). The enzyme can be produced from native or recombinant host cells and can be used to aid in the digestion of cellulose where desired. Also included are other endoglucanases having homology to SEQ ID NO:2.

FIGURE 1A

(SEQ ID NO:1-nucleotide sequence and SEQ ID NO:2-amino acid sequence)

AEPII1a Archaeal Endoglucanase Sequence

```
                9          18          27          36          45          54
5'  ATG ATA AAC GTT GCA ACG GGA GAG GAG ACC CCA ATA CAC CTC TTT GGA GTC AAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Met Ile Asn Val Ala Thr Gly Glu Glu Thr Pro Ile His Leu Phe Gly Val Asn

               63          72          81          90          99         108
    TGG TTC GGC TTT GAG ACA CCG AAC TAC GTT GTT CAC GGC CTA TGG AGT AGG AAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Trp Phe Gly Phe Glu Thr Pro Asn Tyr Val Val His Gly Leu Trp Ser Arg Asn

              117         126         135         144         153         162
    TGG GAG GAC ATG CTC CTC CAG ATC AAG AGC CTT GGC TTC AAT GCG ATA AGG CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Trp Glu Asp Met Leu Leu Gln Ile Lys Ser Leu Gly Phe Asn Ala Ile Arg Leu

              171         180         189         198         207         216
    CCC TTC TGT ACC CAG TCA GTA AAA CCG GGG ACG ATG CCA ACG GCG ATT GAC TAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Phe Cys Thr Gln Ser Val Lys Pro Gly Thr Met Pro Thr Ala Ile Asp Tyr

              225         234         243         252         261         270
    GCC AAG AAC CCA GAC CTC CAG GGT CTT GAC AGC GTC CAG ATA ATG GAG AAA ATA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ala Lys Asn Pro Asp Leu Gln Gly Leu Asp Ser Val Gln Ile Met Glu Lys Ile

              279         288         297         306         315         324
    ATC AAG AAG GCT GGA GAC CTG GGC ATA TTC GTG CTC CTC GAC TAC CAC AGA ATA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ile Lys Lys Ala Gly Asp Leu Gly Ile Phe Val Leu Leu Asp Tyr His Arg Ile
```

```
       333         342         351         360         369         378
GGA TGC AAC TTC ATA GAA CCC CTA TGG TAC ACC GAC AGC TTC TCG GAG CAG GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Cys Asn Phe Ile Glu Pro Leu Trp Tyr Thr Asp Ser Phe Ser Glu Gln Asp

       387         396         405         414         423         432
TAC ATA AAC ACC TGG GTT GAA GTC GCC CAG AGG TTC GGC AAG TAC TGG AAC GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Ile Asn Thr Trp Val Glu Val Ala Gln Arg Phe Gly Lys Tyr Trp Asn Val

       441         450         459         468         477         486
ATC GGC GCG GAC CTG AAG AAC GAA CCC CAC AGC TCA AGC CCC GCA CCT GCC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ile Gly Ala Asp Leu Lys Asn Glu Pro His Ser Ser Ser Pro Ala Pro Ala Ala

       495         504         513         522         531         540
TAC ACT GAC GGA AGT GGG GCC ACG TGG GGA ATG GGC AAC AAC GCC ACC GAC TGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Thr Asp Gly Ser Gly Ala Thr Trp Gly Met Gly Asn Asn Ala Thr Asp Trp

       549         558         567         576         585         594
AAC CTG GCG GCT GAG AGG ATA GGA AGG GCA ATT CTG GAG GTT GCC CCA CAA TGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asn Leu Ala Ala Glu Arg Ile Gly Arg Ala Ile Leu Glu Val Ala Pro Gln Trp

       603         612         621         630         639         648
GTT ATA TTT GTT GAG GGA ACC CAG TTC ACC ACC CCC GAG ATA GAC GGT AGG TAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Ile Phe Val Glu Gly Thr Gln Phe Thr Thr Pro Glu Ile Asp Gly Arg Tyr

       657         666         675         684         693         702
AAG TGG GGC CAC AAC GCC TGG TGG GGC GGA AAC CTT ATG GGT GTT AGG AAG TAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lys Trp Gly His Asn Ala Trp Trp Gly Gly Asn Leu Met Gly Val Arg Lys Tyr
```

```
       711         720         729         738         747         756
CCA GTT AAC CTG CCC AGG GAC AAG GTT GTT TAC AGC CCC CAA GTT TAC GGT TCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Val Asn Leu Pro Arg Asp Lys Val Val Tyr Ser Pro Gln Val Tyr Gly Ser

       765         774         783         792         801         810
GAA GTT TAC GAC CAG CCC TAC TTT GAC CCC GGT GAG GGG TTC CCC GAC AAC CTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Val Tyr Asp Gln Pro Tyr Phe Asp Pro Gly Glu Gly Phe Pro Asp Asn Leu

       819         828         837         846         855         864
CCC GAA ATA TGG TAC CAC CAC TTC GGC TAC GTA AAG CTT GAT CTC GGT TAC CCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Glu Ile Trp Tyr His His Phe Gly Tyr Val Lys Leu Asp Leu Gly Tyr Pro

       873         882         891         900         909         918
GTT GTT ATA GGT GAG TTC GGA GGC AAG TAC GGC CAT GGG GGA GAC CCG AGG GAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Val Ile Gly Glu Phe Gly Gly Lys Tyr Gly His Gly Gly Asp Pro Arg Asp

       927         936         945         954         963         972
GTC ACT TGG CAG AAC AAG ATA ATA GAC TGG ATG ATC CAG AAC AAA TTC TGT GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Thr Trp Gln Asn Lys Ile Ile Asp Trp Met Ile Gln Asn Lys Phe Cys Asp

       981         990         999        1008        1017        1026
TTC TTC TAC TGG AGC TGG AAC CCA AAC AGC GGT GAC ACC GGT GGA ATT CTG AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Phe Phe Tyr Trp Ser Trp Asn Pro Asn Ser Gly Asp Thr Gly Gly Ile Leu Lys

      1035        1044        1053        1062        1071        1080
GAT GAC TGG ACG ACA ATA TGG GAG GAC AAG TAC AAC AAC CTG AAG AGG CTC ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asp Asp Trp Thr Thr Ile Trp Glu Asp Lys Tyr Asn Asn Leu Lys Arg Leu Met
```

```
      1089        1098        1107        1116        1125        1134
GAC AGC TGT TCT GGA AAC GCC ACT GCC CCG TCC GTC CCC ACG ACA ACT ACA ACA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asp Ser Cys Ser Gly Asn Ala Thr Ala Pro Ser Val Pro Thr Thr Thr Thr Thr

      1143        1152        1161        1170        1179        1188
ACA AGC ACA CCG CCA ACG ACC ACA ACG ACT ACA ACA TCC ACT CCA ACG ACC ACT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Ser Thr Pro Pro Thr Thr Thr Thr Thr Thr Thr Ser Thr Pro Thr Thr Thr

      1197        1206        1215        1224        1233        1242
ACC CAG ACC CCG ACC ACC ACT ACT CCA ACT ACG ACA ACC ACC ACG ACC ACA ACT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Gln Thr Pro Thr Thr Thr Thr Pro Thr Thr Thr Thr Thr Thr Thr Thr Thr
      1251        1260        1269        1278        1287        1296
CCT TCA AAT AAC GTC CCA TTT GAA ATT GTG AAC GTT CTC CCG ACT AGC TCC CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ser Asn Asn Val Pro Phe Glu Ile Val Asn Val Leu Pro Thr Ser Ser Gln

      1305        1314        1323        1332        1341        1350
TAC GAG GGA ACC AGC GTG GAG GTT GTA TGT GAT GGA ACC CAG TGT GCC TCC AGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Glu Gly Thr Ser Val Glu Val Val Cys Asp Gly Thr Gln Cys Ala Ser Ser

      1359        1368        1377        1386        1395        1404
GTT TGG GGA GCT CCG AAC CTC TGG GGA GTC GTT AAA ATC GGA AAC GCC ACC ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Trp Gly Ala Pro Asn Leu Trp Gly Val Val Lys Ile Gly Asn Ala Thr Met

      1413        1422        1431        1440        1449        1458
GAC CCC AAC GTT TGG GGC TGG GAG GAC GTT TAC AAG ACT GCA CCC CAG GAC ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asp Pro Asn Val Trp Gly Trp Glu Asp Val Tyr Lys Thr Ala Pro Gln Asp Ile

      1467        1476        1485        1494        1503        1512
```

GGA ACC GGC AGC ACA AAG ATG GAG ATA AGG AAC GGG GTG CTC AAG GTT ACA AAC
Gly Thr Gly Ser Thr Lys Met Glu Ile Arg Asn Gly Val Leu Lys Val Thr Asn

1521 1530 1539 1548 1557 1566
CTC TGG AAC ATC AAC ATG CAT CCG AAG TAT AAC ACA ATG GCA TAC CCG GAG GTC
Leu Trp Asn Ile Asn Met His Pro Lys Tyr Asn Thr Met Ala Tyr Pro Glu Val

1575 1584 1593 1602 1611 1620
ATA TAC GGC GCC AAG CCT TGG GGC AAC CAG CCA ATA AAC GCT CCG AAC TTC GTG
Ile Tyr Gly Ala Lys Pro Trp Gly Asn Gln Pro Ile Asn Ala Pro Asn Phe Val

1629 1638 1647 1656 1665 1674
CTC CCG ATA AAG GTC TCC CAG CTT CCG AGG ATA CTC GTT GAC ACA AAG TAC ACG
Leu Pro Ile Lys Val Ser Gln Leu Pro Arg Ile Leu Val Asp Thr Lys Tyr Thr

1683 1692 1701 1710 1719 1728
CTC GAA AAG AGC TTC CCG GGA AAC AAC TTC GCC TTT GAG GCC TGG CTC TTC AAG
Leu Glu Lys Ser Phe Pro Gly Asn Asn Phe Ala Phe Glu Ala Trp Leu Phe Lys

1737 1746 1755 1764 1773 1782
GAT GCC AAC AAC ATG AGG GCA CCA GGC CAG GGG GAC TAC GAG ATA ATG GTA CAG
Asp Ala Asn Asn Met Arg Ala Pro Gly Gln Gly Asp Tyr Glu Ile Met Val Gln

1791 1800 1809 1818 1827 1836
CTC TAC ATC GAG GGC GGC TAT CCT GCG GGC TAC GAC AAG GGG CCA GTT CTC ACC
Leu Tyr Ile Glu Gly Gly Tyr Pro Ala Gly Tyr Asp Lys Gly Pro Val Leu Thr

1845 1854 1863 1872 1881 1890
GTT GAT GTT CCG ATA ATC GTC GAT GGA AGG CTT GTA AAC CAG ACT TTT GAG CTC

--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Asp Val Pro Ile Ile Val Asp Gly Arg Leu Val Asn Gln Thr Phe Glu Leu

1899 1908 1917 1926 1935 1944
TAC GAC GTC ATA GCG GAT GCC GGA TGG AGG TTC TTC ACC TTC AAG CCA ACT AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Asp Val Ile Ala Asp Ala Gly Trp Arg Phe Phe Thr Phe Lys Pro Thr Lys

1953 1962 1971 1980 1989 1998
AAC TAC AAC GGC TCA GAG GTT GTG TTC GAC TAC ACC AAA TTC ATA GAA ATA GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asn Tyr Asn Gly Ser Glu Val Val Phe Asp Tyr Thr Lys Phe Ile Glu Ile Val

2007 2016 2025 2034 2043 2052
GAC AAC TAC CTC GGC GGT GGC AGC CTC ACG AAC CAC TAC CTG ATG TCC CTG GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asp Asn Tyr Leu Gly Gly Gly Ser Leu Thr Asn His Tyr Leu Met Ser Leu Glu

2061 2070 2079 2088 2097 2106
TTC GGT ACC GAG ATA TAC ACC AAC GGG TGC ACC TCA TTC CCA TGC ACA GTG GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Phe Gly Thr Glu Ile Tyr Thr Asn Gly Cys Thr Ser Phe Pro Cys Thr Val Asp

2115 2124 2133 2142 2151 2160
GTA AGG TGG ACC CTT GAC AAG TAC AGG TTC ATC CTG GCC CCA GGA ACA ATG GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Arg Trp Thr Leu Asp Lys Tyr Arg Phe Ile Leu Ala Pro Gly Thr Met Ala

2169 2178 2187 2196 2205 2214
ACT GAG GAG GCC ATG AGA GTT CTC GTC GGA GAG GTC CAG CCT CCC GCT TCC ACA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Glu Glu Ala Met Arg Val Leu Val Gly Glu Val Gln Pro Pro Ala Ser Thr

2223 2232 2241 2250 2259 2268
ACA ACA TCG CAG ACG ACT ACT TCA ACC ACA ACC CCA ACG CCC ACT ACC ACT ACT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

```
Thr Thr Ser Gln Thr Thr Thr Ser Thr Thr Thr Pro Thr Pro Thr Thr Thr Thr

      2277        2286        2295        2304        2313        2322
ACG ACT CAG ACT TCA ACC ACC ACT ACA ACC ACC TCA CCG CCG ACA ACC ACC GCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Thr Gln Thr Ser Thr Thr Thr Thr Thr Thr Ser Pro Pro Thr Thr Thr Ala

      2331        2340        2349        2358        2367        2376
CCT GCT CAG GAC GTA ATT AAG CTC AGG TAC CCG GAC GAT GGG CAG TGG CCC GAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ala Gln Asp Val Ile Lys Leu Arg Tyr Pro Asp Asp Gly Gln Trp Pro Glu

      2385        2394        2403        2412        2421        2430
GCC CCA ATT GAC AGG GAT GGA GAC GGA AAC CCA GAG TTC TAC ATA GAA ATA AAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Pro Ile Asp Arg Asp Gly Asp Gly Asn Pro Glu Phe Tyr Ile Glu Ile Asn

      2439        2448        2457        2466        2475        2484
CCG TGG AAC ATA CTG AGC GCT GAA AGC TAC GCC GAG ATG ACC TAC AAC TTG AGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Trp Asn Ile Leu Ser Ala Glu Ser Tyr Ala Glu Met Thr Tyr Asn Leu Ser

      2493        2502        2511        2520        2529
AGC GGG GTT CTC CAC TAC GTC CAG GCC CTG GAT AGT ATA TGA TGA 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ser Gly Val Leu His Tyr Val Gln Ala Leu Asp Ser Ile *** ***
```

OC9a (clone # 27GA1) Glycosidase

1
ATG CCA ACC AAT GTA TTT TTC AAC GCC CAT CAC TCG CCG GTT GGG GCG TTT
Met Pro Thr Asn Val Phe Phe Asn Ala His His Ser Pro Val Gly Ala Phe

GCC AGC TTT ACG CTA GGG TTT CCG GGA AAA AGC GGA GGA CTG GAC TTG GAA
Ala Ser Phe Thr Leu Gly Phe Pro Gly Lys Ser Gly Gly Leu Asp Leu Glu

CTT GCC CGA CCG CCA CGG CAA AAT GTC TTT ATT GGC GTT GAG TCG CCG CAT
Leu Ala Arg Pro Pro Arg Gln Asn Val Phe Ile Gly Val Glu Ser Pro His

GAG CCG GGG CTG TAT CAT ATC CTT CCA TTC GCG GAA ACA GCA GGC GAG GAT
Glu Pro Gly Leu Tyr His Ile Leu Pro Phe Ala Glu Thr Ala Gly Glu Asp

GAA AGC AAA CGA TAT GAC ATT GAA AAT CCT GAT CCG AAT CCG CAA AAA CCA
Glu Ser Lys Arg Tyr Asp Ile Glu Asn Pro Asp Pro Asn Pro Gln Lys Pro

AAC ATC CTG ATT CCA TTT GCG AAA GAG CGG ATC GAA CGC GAA TTT CGC GTT
Asn Ile Leu Ile Pro Phe Ala Lys Glu Arg Ile Glu Arg Glu Phe Arg Val

GCC ACG GAT ACA TGG AAG GCC GGG GAC TTG ACG TTG ACG ATT TAT TCA CCG
Ala Thr Asp Thr Trp Lys Ala Gly Asp Leu Thr Leu Thr Ile Tyr Ser Pro

GTG AAG GCC GTA CCA GAT CCG GAA ACG GCC TCC GAG GAA GAA CTC AAG TTG
Val Lys Ala Val Pro Asp Pro Glu Thr Ala Ser Glu Glu Glu Leu Lys Leu

GCG TTG GTT CCA GCT GTC ATT GTC GAG ATG ACG ATC GAT AAT ACG AAC GGA
Ala Leu Val Pro Ala Val Ile Val Glu Met Thr Ile Asp Asn Thr Asn Gly

ACA AGA ACA CGA CGG GCG TTT TTC GGA TTC GAA GGC ACT GAC CCG TAT ACC
Thr Arg Thr Arg Arg Ala Phe Phe Gly Phe Glu Gly Thr Asp Pro Tyr Thr

TCG ATG CGG GGG ATC GAT GAT ACA TGC CCG CAG CTG CGC GGT GTC GGT CAA
Ser Met Arg Gly Ile Asp Asp Thr Cys Pro Gln Leu Arg Gly Val Gly Gln

GGG CGG ATT TTG GGC ATA GCA TCC AAG GAT GAG GGC GTT CGT TCA GCA CTG
Gly Arg Ile Leu Gly Ile Ala Ser Lys Asp Glu Gly Val Arg Ser Ala Leu

CAT TTT AGC ATG GAG GAT ATC TTA ACG GCG ACT CTC GAA GAA AAC TGG ACG
His Phe Ser Met Glu Asp Ile Leu Thr Ala Thr Leu Glu Glu Asn Trp Thr

TTT GGG CTC GGG AAA GTC GGT GCA TTA ATT GCG GAT GTG CCG GCG GGA GAA
Phe Gly Leu Gly Lys Val Gly Ala Leu Ile Ala Asp Val Pro Ala Gly Glu

AAG AAA ACG TAT CAA TTT GCT GTT TGC TTC TAT CGT GGG GGT TGT GTG ACG
Lys Lys Thr Tyr Gln Phe Ala Val Cys Phe Tyr Arg Gly Gly Cys Val Thr

GCG GGA ATG GAT GCC TCT TAT TTT TAC ACC CGT TTC TTC CAT AAT ATC GAA
Ala Gly Met Asp Ala Ser Tyr Phe Tyr Thr Arg Phe Phe His Asn Ile Glu

GAA GTC GGT CTT TAT GCG TTA GAG CAG GCC GAG GTG TTA AAA GAG CAG GCG
Glu Val Gly Leu Tyr Ala Leu Glu Gln Ala Glu Val Leu Lys Glu Gln Ala

TTC CGT TCG AAT GAA CTC ATT GAA AAA GAA TGG CTC TCC GAT GAT CAA AAG
Phe Arg Ser Asn Glu Leu Ile Glu Lys Glu Trp Leu Ser Asp Asp Gln Lys

TTT ATG ATG GCG CAC GCG ATC CGT AGC TAC TAT GGC AAT ACA CAG CTG CTT
Phe Met Met Ala His Ala Ile Arg Ser Tyr Tyr Gly Asn Thr Gln Leu Leu

GAG CAT GAA GGA AAG CCG ATT TGG GTC GTC AAT GAA GGC GAG TAC CGG ATG
Glu His Glu Gly Lys Pro Ile Trp Val Val Asn Glu Gly Glu Tyr Arg Met

ATG AAT ACG TTT GAT CTC ACC GTC GAC CAG CTC TTT TTT GAA TTG AAA ATG
Met Asn Thr Phe Asp Leu Thr Val Asp Gln Leu Phe Phe Glu Leu Lys Met

AAT CCG TGG ACG GTG AAA AAT GTG CTT GAC TTT TAT GTC GAG CGC TAC AGC
Asn Pro Trp Thr Val Lys Asn Val Leu Asp Phe Tyr Val Glu Arg Tyr Ser

TAT GAG GAT CGT GTC CGT TTC CCA GGA GAT GAG ACG GAA TAC CCC GGC GGC

Tyr Glu Asp Arg Val Arg Phe Pro Gly Asp Glu Thr Glu Tyr Pro Gly Gly

ATC AGC TTC ACT CAC GAT ATG GGA GTC GCC AAC ACG TTC TCA CGC CCG CAT

Ile Ser Phe Thr His Asp Met Gly Val Ala Asn Thr Phe Ser Arg Pro His

TAC TCG TCA TAT GAG CTA TAC GGG ATC AGC GGC TGC TTT TCA CAT ATG ACG

Tyr Ser Ser Tyr Glu Leu Tyr Gly Ile Ser Gly Cys Phe Ser His Met Thr

CAC GAA CAG CTC GTC AAC TGG GTG CTT TGC GCA GCG GTA TAC ATC GAA CAA

His Glu Gln Leu Val Asn Trp Val Leu Cys Ala Ala Val Tyr Ile Glu Gln

ACG AAA GAC TGG GCA TGG CGC GAC CGG CGG CTT ACG ATC TTG GAA CAA TGT

Thr Lys Asp Trp Ala Trp Arg Asp Arg Arg Leu Thr Ile Leu Glu Gln Cys

CTC GAA AGC ATG GTG CGC CGC GAT CAT CCG GAT CCA GAA AAG CGG AAC GGC

Leu Glu Ser Met Val Arg Arg Asp His Pro Asp Pro Glu Lys Arg Asn Gly

GTG ATG GGG CTT GAC AGC ACC CGC ACG ATG GGT GGA GCG GAA ATC ACA ACG

Val Met Gly Leu Asp Ser Thr Arg Thr Met Gly Gly Ala Glu Ile Thr Thr

TAT GAT AGT TTG GAT GTT TCT CTT GGC CAG GCG CGC AAC AAT TTA TAT TTG

Tyr Asp Ser Leu Asp Val Ser Leu Gly Gln Ala Arg Asn Asn Leu Tyr Leu

GCA GGA AAA TGT TGG GCT GCC TAT GTG GCG CTC GAA AAG TTG TTC CGC GAT

Ala Gly Lys Cys Trp Ala Ala Tyr Val Ala Leu Glu Lys Leu Phe Arg Asp

GTC GGC AAA GAA GAA CTG GCT GCA TTG GCA AGG GAG CAG GCG GAA AAA TGC

Val Gly Lys Glu Glu Leu Ala Ala Leu Ala Arg Glu Gln Ala Glu Lys Cys

GCC GCG ACG ATT GTC AGT CAC GTG ACG GAG GAC GGG TAT ATC CCA GCC GTG

Ala Ala Thr Ile Val Ser His Val Thr Glu Asp Gly Tyr Ile Pro Ala Val

ATG GGA GAA GGA AAT GAC TCG AAA ATC ATT CCG GCT ATT GAG GGG CTT GTG

Met Gly Glu Gly Asn Asp Ser Lys Ile Ile Pro Ala Ile Glu Gly Leu Val

TTT CCT TAC TTT ACG AAC TGC CAT GAG GCG TTA AGA GAA GAC GGA CGT TTT

Phe Pro Tyr Phe Thr Asn Cys His Glu Ala Leu Arg Glu Asp Gly Arg Phe

GGA GAC TAT ATT CGT GCA CTG CGA CAA CAT TTG CAA TAT GTG TTG CGG GAA

Gly Asp Tyr Ile Arg Ala Leu Arg Gln His Leu Gln Tyr Val Leu Arg Glu

GGA ATT TAC CTA TTC CCG GAC GGG GGA TGG AAA ATT TGC CTC GAC AAG CAA

Gly Ile Tyr Leu Phe Pro Asp Gly Gly Trp Lys Ile Cys Leu Asp Lys Gln

CAA CTC GTG GTT GAG CAA AAT TTA CTT ATG CCA GTT TAT TGC CCG CCG CAT

Gln Leu Val Val Glu Gln Asn Leu Leu Met Pro Val Tyr Cys Pro Pro His

```
TTT AGG GTG GGA ATG GGA TGA        1958
Phe Arg Val Gly Met Gly END
```

Bankia gouldi mix (Clone # 37GP2) Glycosidase

1
ATG TTG AAA AAA CTG GCT TTA GCA GCC GGG ATC GCA GCA GCA ACA CTG GCT
Met Leu Lys Lys Leu Ala Leu Ala Ala Gly Ile Ala Ala Ala Thr Leu Ala

GCA TCC GGT TCC CAT GGG CAG ACG TTC GCG TAC GGC GAA GCT CTG CAA AAA
Ala Ser Gly Ser His Gly Gln Thr Phe Ala Tyr Gly Glu Ala Leu Gln Lys

TCC ATC TAT TTT TAT GAG GCT CAA CAG GCC GGC CCA CTC CCG GAA TGG AAC
Ser Ile Tyr Phe Tyr Glu Ala Gln Gln Ala Gly Pro Leu Pro Glu Trp Asn

CGC GTT GCC TGG CGT GGC GAC TCA GTT CCT GAT GAC GGT GCC GAC GTC GGA
Arg Val Ala Trp Arg Gly Asp Ser Val Pro Asp Asp Gly Ala Asp Val Gly

CTG GAT TTA CGC GGT GGC TGG TTC GAT GCG GGC GAC CAC GTT AAG TTT GGC
Leu Asp Leu Arg Gly Gly Trp Phe Asp Ala Gly Asp His Val Lys Phe Gly

TTT CCA ATG GCC GCG TCA GCG ACA CTC GTC GCC TGG GGA GGC GTC GAT TAC
Phe Pro Met Ala Ala Ser Ala Thr Leu Val Ala Trp Gly Gly Val Asp Tyr

AAA GAC GCG TAC GAA CAG TCG GGG CAA ATG GAA CAT CTG CGC AAC AAC CTG
Lys Asp Ala Tyr Glu Gln Ser Gly Gln Met Glu His Leu Arg Asn Asn Leu

CGC TTC GTC AAT GAC TAC TTT ATC AGC GCG CAC CCC GCT CCG AAC GTG CTT

Arg Phe Val Asn Asp Tyr Phe Ile Ser Ala His Pro Ala Pro Asn Val Leu

TAC GGG CAG GTT GGC GAT GGC AGT GCA GAC CAT ACC TTC TGG GGT CCC GCT

Tyr Gly Gln Val Gly Asp Gly Ser Ala Asp His Thr Phe Trp Gly Pro Ala

GAG GTT CTG CAC CAC AAG ATC CCC GGC TCG CGC ATT TCT ATG AAG ATT GAC

Glu Val Leu His His Lys Ile Pro Gly Ser Arg Ile Ser Met Lys Ile Asp

GAA AGC TGC CCG GGT ACC GAT CTG GCC GCA GAG ACC GCA GCA GCG ATG GCC

Glu Ser Cys Pro Gly Thr Asp Leu Ala Ala Glu Thr Ala Ala Ala Met Ala

GCG TCT GCG ATG GTT TTT CAG GGT GAG GAC GAT GCT TAC GCA GCA ACC CTG

Ala Ser Ala Met Val Phe Gln Gly Glu Asp Asp Ala Tyr Ala Ala Thr Leu

ATC ACT CAC GCC AAA CAG CTG TGG CAA TTT GCT GAT TCA ACC AAA GGC ACA

Ile Thr His Ala Lys Gln Leu Trp Gln Phe Ala Asp Ser Thr Lys Gly Thr

ACC GGT ACA GAT ACA GCC TAT TCC AAT TGC ATA ACA GGT GCA CAG GGC TTT

Thr Gly Thr Asp Thr Ala Tyr Ser Asn Cys Ile Thr Gly Ala Gln Gly Phe

TAT ACG TCG ACG TAT GGC GTT TAC TAC GAT GAA CTT GCC TGG GGT GCT CTC

Tyr Thr Ser Thr Tyr Gly Val Tyr Tyr Asp Glu Leu Ala Trp Gly Ala Leu

TGG TTA TGG CGC GCA ACT GGA GAA GAC TTC TAC CTG GAA CAA GCC AAG CAT

Trp Leu Trp Arg Ala Thr Gly Glu Asp Phe Tyr Leu Glu Gln Ala Lys His

TAC TAC GGT TTG ATG GGC TTT GAA AAC CAG ACG ACA ACT CCG GTA TAT ACC

Tyr Tyr Gly Leu Met Gly Phe Glu Asn Gln Thr Thr Thr Pro Val Tyr Thr

TGG TCG CTT GGC TGG AAC GAT AAA GCG TAT GCC GTT TAT GTA CTT ATG GCC

Trp Ser Leu Gly Trp Asn Asp Lys Ala Tyr Ala Val Tyr Val Leu Met Ala

GCA CTT GTA GGT GAC GAG GTT TAC CAC GCA GAT GCA CAG CGC TAC CTG GAT

Ala Leu Val Gly Asp Glu Val Tyr His Ala Asp Ala Gln Arg Tyr Leu Asp

CAC TGG AGC GTC GGC GAG GGT AAC CGC ACA CCC AAT GGG CTG ATT CTG GTC

His Trp Ser Val Gly Glu Gly Asn Arg Thr Pro Asn Gly Leu Ile Leu Val

GAC TCC TGG GGG GTA AAC CGC TAT GCG GCC AAC GCG GGT TAT CTC GCA CTC

Asp Ser Trp Gly Val Asn Arg Tyr Ala Ala Asn Ala Gly Tyr Leu Ala Leu

TTT TAT GCA GAT GCG ATT GGC AGT GAC CAC CCC CTT TAT GAT CGT TAC CAC

Phe Tyr Ala Asp Ala Ile Gly Ser Asp His Pro Leu Tyr Asp Arg Tyr His

AAT TTT GGT AAG AAG CAG ATC GAT CAT ATC CTG GGC GAC AAC CCT GAC AAC

Asn Phe Gly Lys Lys Gln Ile Asp His Ile Leu Gly Asp Asn Pro Asp Asn

CAA AGC TAC GTC GTC GGC TTT GGC GAT AAT TTC CCA ATC AAT GTT CAC CAC

Gln Ser Tyr Val Val Gly Phe Gly Asp Asn Phe Pro Ile Asn Val His His

CGT GGC TCC CAC GGT TCC TGG TCC GAC AGC ATT TCC AAC CCG GTT AAT CAA

Arg Gly Ser His Gly Ser Trp Ser Asp Ser Ile Ser Asn Pro Val Asn Gln

CGC CAT GTG CTA TAC GGC GCA GTT GCC GGT GGT CCG CAG GGC GAT ACA GGC

Arg His Val Leu Tyr Gly Ala Val Ala Gly Gly Pro Gln Gly Asp Thr Gly

TAT GAA GAA GAC CGC AAT GAC TAT GTG CAG AAT GAG GTC GCA ACA GAC TAC

Tyr Glu Glu Asp Arg Asn Asp Tyr Val Gln Asn Glu Val Ala Thr Asp Tyr

AAC TCA GGC TTC ACC AGT GCC GTC GCT GCA CTT TAT GAT CAC TAT GGT GGC

Asn Ser Gly Phe Thr Ser Ala Val Ala Ala Leu Tyr Asp His Tyr Gly Gly

GCG CCC CTG GCG AAC TTC CCG CCT CCC GAA CCA GAG TCG GTG GAG TAT CTG

Ala Pro Leu Ala Asn Phe Pro Pro Pro Glu Pro Glu Ser Val Glu Tyr Leu

GTG GGG GCC AAG ATC AAT TCC TCT GGC AAC CGC TTC GTG GAA ATG AAA GCC

Val Gly Ala Lys Ile Asn Ser Ser Gly Asn Arg Phe Val Glu Met Lys Ala

GTT ATT CAA AAC CAC AGC ACA ACA CCC GCC CAA GGT AAA GAC GAC CTT TAC

Val Ile Gln Asn His Ser Thr Thr Pro Ala Gln Gly Lys Asp Asp Leu Tyr

ATG CGC TAT TTC TAT GAT CTG AGC GAA GTA TTT GCC GCA GGC TAC AGT TTG

Met Arg Tyr Phe Tyr Asp Leu Ser Glu Val Phe Ala Ala Gly Tyr Ser Leu

AAT GAT CTA ACG GTG GCG TCC GGA TAC AAC CAA GCC TCG GAT GTG AAT GGC

Asn Asp Leu Thr Val Ala Ser Gly Tyr Asn Gln Ala Ser Asp Val Asn Gly

CTG CAA CAT TGG GAT GGC AAC GTC TAC TAT GTG GAA GCC CAG TTC TAT GAC

Leu Gln His Trp Asp Gly Asn Val Tyr Tyr Val Glu Ala Gln Phe Tyr Asp

GAT GTG GTA TTT CCC GGT GGT CAG TCC GCG CAC CGA CGG GAA GTA CAA TTT

Asp Val Val Phe Pro Gly Gly Gln Ser Ala His Arg Arg Glu Val Gln Phe

CGC GTG TCC CTG CCA ACC ACA TCC AAT CTT GCC GAG TGG GAC AAC ACG AAC

Arg Val Ser Leu Pro Thr Thr Ser Asn Leu Ala Glu Trp Asp Asn Thr Asn

GAC CCC TCG TTT GAT CCA AGT TAT TTA ACG GTC GAT AGT AGT CTG ACT TAC

Asp Pro Ser Phe Asp Pro Ser Tyr Leu Thr Val Asp Ser Ser Leu Thr Tyr

GGT ATC GAC GCG CCG AAA ATT CCA CTC TAC GAC GCC AAC GGC CTG CTC TGG

Gly Ile Asp Ala Pro Lys Ile Pro Leu Tyr Asp Ala Asn Gly Leu Leu Trp

GGC GAG GAG CCA CCC CGT GGC GGA ACT TCC TCC AGC TCA TCG TCG AGC AGT

Gly Glu Glu Pro Pro Arg Gly Gly Thr Ser Ser Ser Ser Ser Ser Ser Ser

TCG TCC TCT AGC TCA TCC AGC AGT TCA TCG TCG AGC AGC TCC TCG AGC AGT

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

TCG TCC TCG AGT AAT TCG TCC TCT AGC TCG TCC AGC TCT TCG TCG AAT TCT

Ser Ser Ser Ser Asn Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Asn Ser

TCG TCG TCT AAC AGC AGT TCC TCG TCC AGC TCA AGC TCA TCG AGC AGT TCC

Ser Ser Ser Asn Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

AGT TCG TCG AGT TCG GGC GGC ACC TGT GCG GAC GTG AAC GTA TAC CCC AAC

Ser Ser Ser Ser Ser Gly Gly Thr Cys Ala Asp Val Asn Val Tyr Pro Asn

TGG ACC GCA CGT GAC TGG GCC GGT GGA GTA CCG AAC CAC GCG GAA GCC GGT

Trp Thr Ala Arg Asp Trp Ala Gly Gly Val Pro Asn His Ala Glu Ala Gly

GAT TTG ATG GTT TAC CAA GGT ACT GTC TAC CAA GCT AAT TGG TAC ACC AAC

Asp Leu Met Val Tyr Gln Gly Thr Val Tyr Gln Ala Asn Trp Tyr Thr Asn

AGT GTG CCT GGC AGT GAT GCA TCC TGG ACC AAC CAA GGG TTA TGT GCC GGC

Ser Val Pro Gly Ser Asp Ala Ser Trp Thr Asn Gln Gly Leu Cys Ala Gly

GGC GGA TCC AGC TCC AGC AGC TCA TCA TCC AGC TCA AGC AGC TCT TCG TCC

Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

AGC AGC AGC TCA AGC TCG TCC AGT GGT GCG TCC GGT TCA TCC TCC AGC TCG

Ser Ser Ser Ser Ser Ser Ser Ser Gly Ala Ser Gly Ser Ser Ser Ser Ser

AGC AGT TCG TCC TCG TCA AGT TCG AGC AGC AGC TCT TCG AGT TCG TCT TCT

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

GGT GGC GGC GCC ATG TGT AAC TGG TAT GGC TGG CAA GTA CCT ATT TGT GAA

Gly Gly Gly Ala Met Cys Asn Trp Tyr Gly Trp Gln Val Pro Ile Cys Glu

AAC ACC CCA TCT GGC TGG GGC AAC GAA AAT GGC CAA ACA TGT GTC GGC CCC

Asn Thr Pro Ser Gly Trp Gly Asn Glu Asn Gly Gln Thr Cys Val Gly Pro

GAT ACT TGC CAA GAG GTC GTC AAC TAA 2627

Asp Thr Cys Gln Glu Val Val Asn END

Bankia gouldi mix (Clone # 37GP3) Glycosidase

1
ATG AAG ATG ACC TAC ATG CAT CCG GCT GAA GAT ACT TAC TCG TTT GGT CAA
Met Lys Met Thr Tyr Met His Pro Ala Glu Asp Thr Tyr Ser Phe Gly Gln

GCG GAT CAG TTG GTC AAC TGG GCG AAA GCG AAT GGT ATT GGC GTG CAC GGC
Ala Asp Gln Leu Val Asn Trp Ala Lys Ala Asn Gly Ile Gly Val His Gly

CAC ACT CTG GTT TGG CAC TCC GAA TAC CAG GTA CCC AAT TGG ATG AAA AAT
His Thr Leu Val Trp His Ser Glu Tyr Gln Val Pro Asn Trp Met Lys Asn

TAC TCT GGT GAT GCA ACT GCA TTC CAA ACC ATG CTC AAC ACC CAT GTG AAA
Tyr Ser Gly Asp Ala Thr Ala Phe Gln Thr Met Leu Asn Thr His Val Lys

ACT GTG GCT GAG CAT TTT GCT GGC GAA CTG GAC AGC TGG GAC GTT GTG AAT
Thr Val Ala Glu His Phe Ala Gly Glu Leu Asp Ser Trp Asp Val Val Asn

GAA GTG CTG GAG CCG GGC TCC AAT GGT TGC TGG CGT GAA AAC TCT CTG TTC
Glu Val Leu Glu Pro Gly Ser Asn Gly Cys Trp Arg Glu Asn Ser Leu Phe

TAC CAG AAG CTT GGC AAA GAC TTT GTC GCG AAC GCA TTC CGT GCA GCT CGC
Tyr Gln Lys Leu Gly Lys Asp Phe Val Ala Asn Ala Phe Arg Ala Ala Arg

GAG GGC GAT CCC AAT GCA GAC TTG TAT TAC AAC GAT TAC TCG ACT GAA AAT
Glu Gly Asp Pro Asn Ala Asp Leu Tyr Tyr Asn Asp Tyr Ser Thr Glu Asn

GGT GTA ACT TCC GAT GAG AAG TTC AGT TGT TTG TTG GAA CTA GTC GAT GAG
Gly Val Thr Ser Asp Glu Lys Phe Ser Cys Leu Leu Glu Leu Val Asp Glu

CTT CTG GAA GCG GAC GTG CCG ATT ACA GGT GTT GGT TTC CAA ATG CAC GTG
Leu Leu Glu Ala Asp Val Pro Ile Thr Gly Val Gly Phe Gln Met His Val

CAG GCG ACG TGG CCT AGC AAT GCC AAC ATC GGC AAG GCA TTC AAA GCC ATC
Gln Ala Thr Trp Pro Ser Asn Ala Asn Ile Gly Lys Ala Phe Lys Ala Ile

GCG GAT CGC GGT CTG AAA GTT AAA ATT TCT GAG CTC GAT GTT CCT GTT AAC
Ala Asp Arg Gly Leu Lys Val Lys Ile Ser Glu Leu Asp Val Pro Val Asn

AAC CCT TAC GGA ACC ACT AAT TTC CCG CAA TAC AGC AGT TTT ACC GCG GAA
Asn Pro Tyr Gly Thr Thr Asn Phe Pro Gln Tyr Ser Ser Phe Thr Ala Glu

GCC GCC GAG CTG CAG AAG CAG CGC TAC AAG GGC ATT ATG CAA GCG TAC CTT
Ala Ala Glu Leu Gln Lys Gln Arg Tyr Lys Gly Ile Met Gln Ala Tyr Leu

GAT AAC GTA CCG GCC AAC CTG CGT GGT GGT TTC ACC GTG TGG GGC GTT TGG
Asp Asn Val Pro Ala Asn Leu Arg Gly Gly Phe Thr Val Trp Gly Val Trp

GAT GGC GAT AGC TGG ATC ATG ACG TTC AGC CAG TAC ACC AAC GCT AAC GCC

Asp Gly Asp Ser Trp Ile Met Thr Phe Ser Gln Tyr Thr Asn Ala Asn Ala

AAC GAC TGG CCA CTG TTG TTC ACG GGG CCG TAA 848

Asn Asp Trp Pro Leu Leu Phe Thr Gly Pro END

Teredinibacter, pure (Clone 42GP1) Glycosidase

1

ATG GGA ACA TCT CTT ATG ATC AAA TCT ACA CTG ACA GGT ATG ATT ACT GCT

Met Gly Thr Ser Leu Met Ile Lys Ser Thr Leu Thr Gly Met Ile Thr Ala

GTT GCC GCC GCA GTT TTC ACC ACC TCT GCA GCT TTC GCG GAT GTA CCT CCG

Val Ala Ala Ala Val Phe Thr Thr Ser Ala Ala Phe Ala Asp Val Pro Pro

TTG ACA GTG AGC GGA AAT CAG GTT TTA AGT GGC GGT GAA GCA AAA AGC TTC

Leu Thr Val Ser Gly Asn Gln Val Leu Ser Gly Gly Glu Ala Lys Ser Phe

GCT GGT AAC AGC TTC TTT TGG AGC AAT ACC GGA TGG GGC CAG GAA CGT TTT

Ala Gly Asn Ser Phe Phe Trp Ser Asn Thr Gly Trp Gly Gln Glu Arg Phe

TAC AAC GCA GAA ACT GTG CGT TGG TTG AAA GAC GAC TGG AAC GCA ACC ATT

Tyr Asn Ala Glu Thr Val Arg Trp Leu Lys Asp Asp Trp Asn Ala Thr Ile

GTC CGC GCC GCT ATG GGC GTA GAC TTT GAT GGC AGC TAT ATC CCC GAG CAT

Val Arg Ala Ala Met Gly Val Asp Phe Asp Gly Ser Tyr Ile Pro Glu His

GAA GAC GCC GAC CCC GAG GGT AAC GTC GCT CGC GTA CGT GCA TTG GTG GAT

Glu Asp Ala Asp Pro Glu Gly Asn Val Ala Arg Val Arg Ala Leu Val Asp

GCA GCC ATC GCA GAA GAC ATG TAC GTG ATT ATC GAT TTT CAC ACT CAC CAC

Ala Ala Ile Ala Glu Asp Met Tyr Val Ile Ile Asp Phe His Thr His His

GCA GAA GAT TAC CAA GCC GAA TCT ATC GAG TTC TTC GAA GAA ATG GCC ACA

Ala Glu Asp Tyr Gln Ala Glu Ser Ile Glu Phe Phe Glu Glu Met Ala Thr

CTG TAC GGT GGG TAC GAC AAT GTT ATT TAT GAA ATC TAT AAC GAG CCC CTG

Leu Tyr Gly Gly Tyr Asp Asn Val Ile Tyr Glu Ile Tyr Asn Glu Pro Leu

CAA ATC AGC TGG GAC AAT GTT ATT AAA CCT TAT GCA GAA TCG GTG ATT GGC

Gln Ile Ser Trp Asp Asn Val Ile Lys Pro Tyr Ala Glu Ser Val Ile Gly

GCT ATC CGC GCA ATC GAC CCG GAC AAC CTG ATT ATC GTC GGC ACG CCC ACT

Ala Ile Arg Ala Ile Asp Pro Asp Asn Leu Ile Ile Val Gly Thr Pro Thr

TGG TCA CAG GAC GTG GAC GCC GCT GCG CGC AAT CCA ATC ACC AGC TAC AGC

Trp Ser Gln Asp Val Asp Ala Ala Ala Arg Asn Pro Ile Thr Ser Tyr Ser

AAT ATT GCG TAC ACC CTG CAC TTT TAC GCA GGC ACT CAC GGT TCA TGG TTG

Asn Ile Ala Tyr Thr Leu His Phe Tyr Ala Gly Thr His Gly Ser Trp Leu

CGC GAT AAA GCG CGT AAC GCT ATG AAC AGT GGT ATT GCG CTG TTT GTG ACT

Arg Asp Lys Ala Arg Asn Ala Met Asn Ser Gly Ile Ala Leu Phe Val Thr

GAG TGG GGC ACC GTT AAT GCA GAT GGC GAT GGT GCG CCT GCA GTT AAC GAA

Glu Trp Gly Thr Val Asn Ala Asp Gly Asp Gly Ala Pro Ala Val Asn Glu

ACT CAG CAA TGG ATG GAC TTC CTC AAG CAG AAC AAT ATC TCT CAC TTG AAC

Thr Gln Gln Trp Met Asp Phe Leu Lys Gln Asn Asn Ile Ser His Leu Asn

TGG TCC GTG AGT GAT AAA TTG GAA GGT GCG TCT ATC GTA CAA CCT GGC ACG

Trp Ser Val Ser Asp Lys Leu Glu Gly Ala Ser Ile Val Gln Pro Gly Thr

CCC ATT AGC GGC TGG AAC GCT TCT GAC CTT ACG GCC TCC GGC ACA CTG GTT

Pro Ile Ser Gly Trp Asn Ala Ser Asp Leu Thr Ala Ser Gly Thr Leu Val

AAG AAC ATC GTT TCC AAC TGG GGC ACC ACA ATC GGT AAC GGC AGC TCC TCA

Lys Asn Ile Val Ser Asn Trp Gly Thr Thr Ile Gly Asn Gly Ser Ser Ser

AGT TCA TCC AGC TCC TCT TCC AGC TCT TCA AGC AGT TCT TCT TCG AGC AGT

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

TCC TCC TCC AGC AGC TCT TCC TCG TCA AGC AGC TCC GGA TCA ACT GGT GGC

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Gly Ser Thr Gly Gly

GGC AAC TGT GCT GGA GTG AAT GTG TAC CCG AAC TGG ACC GCG CGT GAC TGG

Gly Asn Cys Ala Gly Val Asn Val Tyr Pro Asn Trp Thr Ala Arg Asp Trp

TCT GGC GGC GCC TAC AAC CAT GCG AAC GCT GGC GAC CAA ATG GTC TAT CAA
Ser Gly Gly Ala Tyr Asn His Ala Asn Ala Gly Asp Gln Met Val Tyr Gln

AAC AGC CTG TAT CGT GCC AAC TGG TAC ACC AAC AGC GTG CCT GGC AGC GAC
Asn Ser Leu Tyr Arg Ala Asn Trp Tyr Thr Asn Ser Val Pro Gly Ser Asp

GCC TCC TGG ACT AGC CTT GGC GCC TGC GGA GGC AAC GGA AGT ACG ACC TCA
Ala Ser Trp Thr Ser Leu Gly Ala Cys Gly Gly Asn Gly Ser Thr Thr Ser

TCC AGC TCA AGC AGC TCC TCG TCA AGC AGC AGC TCT TCT TCC AGC AGC TCC
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

TCG TCT ACT GGC GGT GGC TCC AGC TCC TCC AGC AGT TCA TCT TCT TCA TCG
Ser Ser Thr Gly Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

TCG TCT TCC AGC AGC TCT AGC AGC ACT GGT GGC GGT CAA TGT ACC GAA GTG
Ser Ser Ser Ser Ser Ser Ser Ser Thr Gly Gly Gly Gln Cys Thr Glu Val

TGC AAC TGG TAC GGT CAG GGA ACC TAC CCA CTG TGT AAC AAC ACC AGT GGT
Cys Asn Trp Tyr Gly Gln Gly Thr Tyr Pro Leu Cys Asn Asn Thr Ser Gly

TGG GGT TGG GAA AAC AAT CAG AGC TGT ATC GGC CGT CAA ACC TGT GAG TCA

Trp Gly Trp Glu Asn Asn Gln Ser Cys Ile Gly Arg Gln Thr Cys Glu Ser

CAG AAC GGT GGC GCT GGC GGC GTG GTG AGC AAC TGC ACC GGT TCG AGT ACA

Gln Asn Gly Gly Ala Gly Gly Val Val Ser Asn Cys Thr Gly Ser Ser Thr

TCC AGC AGC TCC TCT TCC AGC AGT AGT TCT TCC TCA AGT AGC AGC TCC AGT

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser

TCA TCC AGC AGC TCT TCA TCT GGC ACT GGT AGC AGT ACA TCT TCC AGC AGC

Ser Ser Ser Ser Ser Ser Ser Gly Thr Gly Ser Ser Thr Ser Ser Ser Ser

AGC TCT TCC AGC AGC TCC AGC TCA AGT ACC GGT TCC TCC GGT ATG CCT GGA

Ser Ser Ser Ser Ser Ser Ser Ser Ser Thr Gly Ser Ser Gly Met Pro Gly

CCA CGC GTG GAC AAC CCC TTC GCC GCT GCG CAG AAG TGG TAC ATA AAC CCA

Pro Arg Val Asp Asn Pro Phe Ala Ala Ala Gln Lys Trp Tyr Ile Asn Pro

ATG TGG TCA GCG AGT GCT GCA AAC GAA CCC GGC GGC TCT GTC ATT GCC AAC

Met Trp Ser Ala Ser Ala Ala Asn Glu Pro Gly Gly Ser Val Ile Ala Asn

GAA CCC TCG TTT GTA TGG ATG GAC CGT ATC GGC GCA ATC GAA GGG CCT GCT

Glu Pro Ser Phe Val Trp Met Asp Arg Ile Gly Ala Ile Glu Gly Pro Ala

GAC GGT ATG GGC CTG CGC GAC CAC TTG AAC GAA GCC CTT GCA CAA GGC GCC

Asp Gly Met Gly Leu Arg Asp His Leu Asn Glu Ala Leu Ala Gln Gly Ala

GAC CTG TTC ATG TTT GTT GTG TAC GAC CTG CCA AAC CGT GAC TGT GCT GCA

Asp Leu Phe Met Phe Val Val Tyr Asp Leu Pro Asn Arg Asp Cys Ala Ala

CTC GCC TCC AAC GGT GAA CTG CGC ATC TCC GAA GAT GGC TTC AAC ATC TAC

Leu Ala Ser Asn Gly Glu Leu Arg Ile Ser Glu Asp Gly Phe Asn Ile Tyr

AAG TCC GAC TAC ATC GCA CCT ATC GTT GAA ATC ATC AGC GAC CCT GCA TAC

Lys Ser Asp Tyr Ile Ala Pro Ile Val Glu Ile Ile Ser Asp Pro Ala Tyr

GCA GGT ATC AAA ATC GCT GCG GTT ATC GAG GTG GAC TCA CTG CCT AAC CTG

Ala Gly Ile Lys Ile Ala Ala Val Ile Glu Val Asp Ser Leu Pro Asn Leu

GTT ACC AAT CTG AGC GAA CCT GAC TGT CAG GAA GCA AAT GGT CCT GGC GGC

Val Thr Asn Leu Ser Glu Pro Asp Cys Gln Glu Ala Asn Gly Pro Gly Gly

TAC CGC GAC GGC ATT CGT CAC GCT ATC ACT GAA CTG GGC AAA ATC CCC AAC

Tyr Arg Asp Gly Ile Arg His Ala Ile Thr Glu Leu Gly Lys Ile Pro Asn

GTA TAC TCC TAC GTG GAT ATT GCA CAC TCA GGC TGG CTG GGC TGG AAC GAC

Val Tyr Ser Tyr Val Asp Ile Ala His Ser Gly Trp Leu Gly Trp Asn Asp

AAC TTC GCG CAA GGC GTT AAC CTG ATT TAT GAA GTG GTT GCC AAC CTC GGT

Asn Phe Ala Gln Gly Val Asn Leu Ile Tyr Glu Val Val Ala Asn Leu Gly

TCC GGC ATT AAC CCA ATC GCC GGT TTC GTC AGT AAC TCC GCT AAC TAC ACG

Ser Gly Ile Asn Pro Ile Ala Gly Phe Val Ser Asn Ser Ala Asn Tyr Thr

CCT GTG GAA GAA CCC TTC TTG CCA GAC GCC AAC CTG CAG GTC GGT GGT CAG

Pro Val Glu Glu Pro Phe Leu Pro Asp Ala Asn Leu Gln Val Gly Gly Gln

CCC GTT CGC TCT TCC GAT TTC TAT GAG TGG AAC AGC TAC CTG GCA GAG AAA

Pro Val Arg Ser Ser Asp Phe Tyr Glu Trp Asn Ser Tyr Leu Ala Glu Lys

CCC TTC GTG ACC GAT TGG CGT TCT GCC ATG ATC TCG AAA GGT ATG CCA AGC

Pro Phe Val Thr Asp Trp Arg Ser Ala Met Ile Ser Lys Gly Met Pro Ser

TCC ATC GGT ATG CTG ATC GAT ACC GCA CGT AAC GGC TGG GGT GGC CCT GAG

Ser Ile Gly Met Leu Ile Asp Thr Ala Arg Asn Gly Trp Gly Gly Pro Glu

CGT CCA ACT GCG CAG TCT ACC TCC AAC AAC CTG AAC ACC TTC GTT AAC GAA

Arg Pro Thr Ala Gln Ser Thr Ser Asn Asn Leu Asn Thr Phe Val Asn Glu

TCA CGT ATC GAC CGT CGT GAG CAC CGC GGC AAC TGG TGT AAC CAG CCT GGT
Ser Arg Ile Asp Arg Arg Glu His Arg Gly Asn Trp Cys Asn Gln Pro Gly

GGT GTC GGC TAC CGT CCA ACC GCT GCA CCT TCT CCA GGT ATT GAT GCC TAC
Gly Val Gly Tyr Arg Pro Thr Ala Ala Pro Ser Pro Gly Ile Asp Ala Tyr

GTT TGG GTG AAA CCA CAG GGT GAG TCT GAC GGT GTT TCC GAT CCT AAC TTC
Val Trp Val Lys Pro Gln Gly Glu Ser Asp Gly Val Ser Asp Pro Asn Phe

GAG ATC GAT CCT AAC GAC CCG AAC AAA CAG CAC GAC CCA ATG TGT GAT CCG
Glu Ile Asp Pro Asn Asp Pro Asn Lys Gln His Asp Pro Met Cys Asp Pro

TTC GCC AGC AAC TCG TCC AAC AGT GCA TAC GGC ACC GGC GCT ATG CCA AAT
Phe Ala Ser Asn Ser Ser Asn Ser Ala Tyr Gly Thr Gly Ala Met Pro Asn

GCT CCG CAC GCT GGT CGC TGG TTC CCT GAA GCC TTC CAG TTA CTG CTT GAA
Ala Pro His Ala Gly Arg Trp Phe Pro Glu Ala Phe Gln Leu Leu Leu Glu

AAC GCT TAC CCA CCA ATT AAC TAA 3032
Asn Ala Tyr Pro Pro Ile Asn END

Microscilla furvescens (Clone # 53GC1)

1
ATG AAC AAG AAG TGG TGG AAA GAA GCC GTG GTG TAT CAA GTC TAC CCG CGG

Met Asn Lys Lys Trp Trp Lys Glu Ala Val Val Tyr Gln Val Tyr Pro Arg

AGC TTC AAA GAC AGC AAT GGA GAT GGT GTA GGC GAT CTG CCT GGG GTT ATT

Ser Phe Lys Asp Ser Asn Gly Asp Gly Val Gly Asp Leu Pro Gly Val Ile

GAA AAG CTT GAT TAC ATC AAA AGC CTT GGG GTG GAT GTT ATC TGG CTA TGC

Glu Lys Leu Asp Tyr Ile Lys Ser Leu Gly Val Asp Val Ile Trp Leu Cys

CCG GTG TAC GAT TCC CCC AAT GAT GAC AAT GGT TAC GAT ATT CGT GAC TAC

Pro Val Tyr Asp Ser Pro Asn Asp Asp Asn Gly Tyr Asp Ile Arg Asp Tyr

TAC GAT ATC ATG GCT GAT TTC GGC ACG ATG GCT GAT TTT GAT CAG CTG CTC

Tyr Asp Ile Met Ala Asp Phe Gly Thr Met Ala Asp Phe Asp Gln Leu Leu

GAG GGA ATA CAT CAG CGT GGG ATG AAA CTG CTA ATG GAC CTG GTG GTA AAC

Glu Gly Ile His Gln Arg Gly Met Lys Leu Leu Met Asp Leu Val Val Asn

CAC TGC TCT GAT GAG CAC AAA TGG TTT CAG GAG TCC CGC AAG AGT AAA GAC

His Cys Ser Asp Glu His Lys Trp Phe Gln Glu Ser Arg Lys Ser Lys Asp

AAC CCT TAC CGG GAC TAC TTC ATC TGG AAG CCT GGC AAA AAC GGA GGC CCA

Asn Pro Tyr Arg Asp Tyr Phe Ile Trp Lys Pro Gly Lys Asn Gly Gly Pro

CCT AAC AAC TGG CAG TCC TTT TTT AGT GGT AAT GCC TGG GAA TAC GAT GAG

Pro Asn Asn Trp Gln Ser Phe Phe Ser Gly Asn Ala Trp Glu Tyr Asp Glu

GCC ACT GAC GAG TAT TAC CTA CAT CTT TTC ACC AAA AAG CAA CCA GAC CTC

Ala Thr Asp Glu Tyr Tyr Leu His Leu Phe Thr Lys Lys Gln Pro Asp Leu

AAT TGG GAA AAC CCG AAA GTA CGT GAG GAG GTG CAC AAG CTG ATG AAG TAT

Asn Trp Glu Asn Pro Lys Val Arg Glu Glu Val His Lys Leu Met Lys Tyr

TGG CTG GAC AAA GGA GTA GAT GGG TTC CGG ATG GAT GTG ATT TCC GTG ATT

Trp Leu Asp Lys Gly Val Asp Gly Phe Arg Met Asp Val Ile Ser Val Ile

TCA AAA AGA AAC TTC GAA GAT TCA CCT TAC AAG GAC TTC AAC AAG ACC ATC

Ser Lys Arg Asn Phe Glu Asp Ser Pro Tyr Lys Asp Phe Asn Lys Thr Ile

GAT AAC GTC TAC GCC AAT GGC CCG CGT GTG CAG GAG TTT CTC CAG GAA ATG

Asp Asn Val Tyr Ala Asn Gly Pro Arg Val Gln Glu Phe Leu Gln Glu Met

AAC CGT GAA GTA CTG AGT AAG TAC GAT GTG ATG ACA GTA GGT GAG GGT CCA

Asn Arg Glu Val Leu Ser Lys Tyr Asp Val Met Thr Val Gly Glu Gly Pro

GGT ATC AAT CTG GAA AGC GGC CTG CAA TAT GTA TCC AGC TCA GCG GAG GCT
Gly Ile Asn Leu Glu Ser Gly Leu Gln Tyr Val Ser Ser Ser Ala Glu Ala

CTT AAT ATG ATT TTT CAT TTT GGG CAC ATG TTT ATG GAT CAT GGA CCC GGA
Leu Asn Met Ile Phe His Phe Gly His Met Phe Met Asp His Gly Pro Gly

GGT AGA TTT GAT CCC AAG CCC ATC GAT TTT CTG GAA TTC AAA AAA GTC TTC
Gly Arg Phe Asp Pro Lys Pro Ile Asp Phe Leu Glu Phe Lys Lys Val Phe

AGG CTG TGG GAT GAG TAC CTT AAA GAA GAG GGC TGG GGT AGC GTC TTT CTA
Arg Leu Trp Asp Glu Tyr Leu Lys Glu Glu Gly Trp Gly Ser Val Phe Leu

GGG AAT CAT GAT TTT CAG CGA ATC GTT TCT CGC TTT GGG GAT GAC GGA GCG
Gly Asn His Asp Phe Gln Arg Ile Val Ser Arg Phe Gly Asp Asp Gly Ala

TAC TGG AAA GAG TCC GCC AAA CTG CTG AGC TTG TTG CTA TTT AGC ATG CGC
Tyr Trp Lys Glu Ser Ala Lys Leu Leu Ser Leu Leu Leu Phe Ser Met Arg

GGC ACG GTC TAC GTT TAC CAG GGT GAT GAA ATA GGT ATG ACC AAT GTG GCT
Gly Thr Val Tyr Val Tyr Gln Gly Asp Glu Ile Gly Met Thr Asn Val Ala

TTT GAC ACC ATA GAA GAA TAT GAC GAT GTG GAG ATC AAA AAT GCT TAC AAG

Phe Asp Thr Ile Glu Glu Tyr Asp Asp Val Glu Ile Lys Asn Ala Tyr Lys

GAG TGG AAA GCT GAA GGA AAA GAC CTG GAT CAG TTT TTA AAG AAC GTC CAT

Glu Trp Lys Ala Glu Gly Lys Asp Leu Asp Gln Phe Leu Lys Asn Val His

ATC AAT GGC CGT GAC AAT GCC CGT ACA CCG CTG CAA TGG AAT GAT GCT GAG

Ile Asn Gly Arg Asp Asn Ala Arg Thr Pro Leu Gln Trp Asn Asp Ala Glu

CAG GCT GGT TTT ACC TCA GGC ACT CCA TGG CTC AAA GTC AAC CCT AAC TAT

Gln Ala Gly Phe Thr Ser Gly Thr Pro Trp Leu Lys Val Asn Pro Asn Tyr

ACG GCA ATC AAT GTG GCT AGT CAG GAA GGA GAT GAG AAC TCT ATT CTG GCA

Thr Ala Ile Asn Val Ala Ser Gln Glu Gly Asp Glu Asn Ser Ile Leu Ala

TTT TAT CGC CGG ATG GTG GCG ATG CGA AAG GAG CAC CCG ACA CTT GTT TAT

Phe Tyr Arg Arg Met Val Ala Met Arg Lys Glu His Pro Thr Leu Val Tyr

GGT GAT TTT GCC CCC ATT CAG GAA GAT CAT CCG AGT GTA TTT GCT TTT TGG

Gly Asp Phe Ala Pro Ile Gln Glu Asp His Pro Ser Val Phe Ala Phe Trp

AGA TGG GAT GAA GAG GCT GCA TAT TTA GTC TTA CTC AAT TTT TCT GAG GAG

Arg Trp Asp Glu Glu Ala Ala Tyr Leu Val Leu Leu Asn Phe Ser Glu Glu

ACT CAG GAA TTT GGG CTG GAC GAT CGA TTT GAT AGT AGT AAG CTT CGC ATA

Thr Gln Glu Phe Gly Leu Asp Asp Arg Phe Asp Ser Ser Lys Leu Arg Ile

GTA GAG GCC AAT GAC TTT GAC TTT GGT GAG CCA CAA AGT GGA AAA GTG AAA

Val Glu Ala Asn Asp Phe Asp Phe Gly Glu Pro Gln Ser Gly Lys Val Lys

1682

CTA AAA CCG TGG CAG GCG GTG TTG GCG CGT GTT CGG CAT ATT GAA TTG TAA
Leu Lys Pro Trp Gln Ala Val Leu Ala Arg Val Arg His Ile Glu Leu END

Thermotoga neapolitana (Clone #56GC2) Glycosidase

1

TCT TCT GAA CGA TTC TCC ACT GAG CAG AAA AGA CCA GAT CAT ACT CTT TGT

Ser Ser Glu Arg Phe Ser Thr Glu Gln Lys Arg Pro Asp His Thr Leu Cys

GGA CGG AAA AGA ACA TTC GGC AAA GAA GGT GGT TAT ACC ACC CTT CAA AGA

Gly Arg Lys Arg Thr Phe Gly Lys Glu Gly Gly Tyr Thr Thr Leu Gln Arg

GGA AAC GCT GGT CTT CAA AGT GAA CGG ACT GAA GAG GGG AGA GCA CCT CGT

Gly Asn Ala Gly Leu Gln Ser Glu Arg Thr Glu Glu Gly Arg Ala Pro Arg

ATC CAC CAG TCT GAA CAC GGG AAA AAC CAT CTA TGT GAG GTG ATC TGT GTG

Ile His Gln Ser Glu His Gly Lys Asn His Leu Cys Glu Val Ile Cys Val

GAG ATC TTC AAA AGA CCG TTC AGA GAA GGG AGC TTC GTT CTG AAA GAG AAG

Glu Ile Phe Lys Arg Pro Phe Arg Glu Gly Ser Phe Val Leu Lys Glu Lys

GAC TAC ACC GTT GAG TTC GAG GTG GAG AAG ATC CAT CTT GGA TGG AAG ATT

Asp Tyr Thr Val Glu Phe Glu Val Glu Lys Ile His Leu Gly Trp Lys Ile

TCA GGG AGA GTG AAG GGA AAT CCC GGA AGG CTT GAG ATC TTT CGG ACA AAC

Ser Gly Arg Val Lys Gly Asn Pro Gly Arg Leu Glu Ile Phe Arg Thr Asn

GCA CCG AAG AAA CTC CTC GTG AAC AAC TGG CAG TCC TGG GGA CCC TGC AGG

Ala Pro Lys Lys Leu Leu Val Asn Asn Trp Gln Ser Trp Gly Pro Cys Arg

GTG GTG GAT CTT CCA TCC TTC ACC CCA CCC GAG ATA GAT CCA AAC TGG CAG

Val Val Asp Leu Pro Ser Phe Thr Pro Pro Glu Ile Asp Pro Asn Trp Gln

TAC ACG GCC TCT GTG GTA CCG GAT GTG ATC AAA AAC CGT CTT CAG AGT GAC

Tyr Thr Ala Ser Val Val Pro Asp Val Ile Lys Asn Arg Leu Gln Ser Asp

TAC TTC GTG GCA GAG GAA GGG AGA GTA TAC GGT TTT TTG AGT TCG AAG ATC

Tyr Phe Val Ala Glu Glu Gly Arg Val Tyr Gly Phe Leu Ser Ser Lys Ile

GCA CAT CCT TTC TTT GCG GCA GAG AAT GGA GAA CTT GTT GCG TAT CTT GAG

Ala His Pro Phe Phe Ala Ala Glu Asn Gly Glu Leu Val Ala Tyr Leu Glu

TAC TTC GAT GTG AAT TTC GAT GAC TTC GTC CCG ATA GAA CCT TTT GTC GTC

Tyr Phe Asp Val Asn Phe Asp Asp Phe Val Pro Ile Glu Pro Phe Val Val

CTT GAA AAT CCA ATC ACC TCT CTC CTT CTG GAA AAG TAC GCT GAA CTC GTC

Leu Glu Asn Pro Ile Thr Ser Leu Leu Leu Glu Lys Tyr Ala Glu Leu Val

GGG AAG GAA AAC AGC GCG AGG ATT CCA AAA CGT ACA CCG GTT GGA TGG TGC

Gly Lys Glu Asn Ser Ala Arg Ile Pro Lys Arg Thr Pro Val Gly Trp Cys

AGC TGG TAC CAC TAT TTC CTC GAT CTC ACC TGG GAG GAG ACT TTG AAG AAT
Ser Trp Tyr His Tyr Phe Leu Asp Leu Thr Trp Glu Glu Thr Leu Lys Asn

CTG GAA CTT GCA GGA GAG TTT CCC TTC GAG GTC TTT CAG ATA GAC GAC GCG
Leu Glu Leu Ala Gly Glu Phe Pro Phe Glu Val Phe Gln Ile Asp Asp Ala

TAT GAA AAA GAC ATC GGA GAC TGG CTC GTC ACG AAG AAA GAC TTC CCA TCT
Tyr Glu Lys Asp Ile Gly Asp Trp Leu Val Thr Lys Lys Asp Phe Pro Ser

GTG GAC GAG ATG GCA AGG ACG ATA CAG GAG AAA GGC TTT GTT CCT GGT ATA
Val Asp Glu Met Ala Arg Thr Ile Gln Glu Lys Gly Phe Val Pro Gly Ile

TGG ACC GCA CCG TTC AGT GTT TCA GAA ACA TCG GAT GTG TTC AAC TCC TAT
Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser Asp Val Phe Asn Ser Tyr

CCG GAC TGG GTC GTG AAG GAA AAC GGA ATG CCA AAG ATG GCG TAC AGG AAC
Pro Asp Trp Val Val Lys Glu Asn Gly Met Pro Lys Met Ala Tyr Arg Asn

TGG AAC AGA AAG ATC TAC GCT CTT GAC CTT TCA AAC AAA GAA GTC CTG GAC
Trp Asn Arg Lys Ile Tyr Ala Leu Asp Leu Ser Asn Lys Glu Val Leu Asp

TGG CTC TTC GAC CTC TTC AGC TCT CTC AAG AAG ATG GGC TAC AGA TAC TTC

Trp Leu Phe Asp Leu Phe Ser Ser Leu Lys Lys Met Gly Tyr Arg Tyr Phe

AAG ATC GAC TTT CTC TTT GCA GGA GCG ATT CCG GGT GAG AGG AAA GAA AAC

Lys Ile Asp Phe Leu Phe Ala Gly Ala Ile Pro Gly Glu Arg Lys Glu Asn

ATC ACA CCC GTT CAG GCG TTC AGA AAG GGG ATG GAG GTG ATC AGA AAG GCG

Ile Thr Pro Val Gln Ala Phe Arg Lys Gly Met Glu Val Ile Arg Lys Ala

GTT GGA GAC TTG TTC ATA CTC GGA TGT GGC TCT CCC CTT CTT CCT GCG GTG

Val Gly Asp Leu Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala Val

GGC TAC GTT GAC GGC ATG AGG ATA GGG CCG GAC ACC ACA CCC TTC TGG GGT

Gly Tyr Val Asp Gly Met Arg Ile Gly Pro Asp Thr Thr Pro Phe Trp Gly

GAT CAA ATA GAA GAC AAC GGA GCA CCC GCT GCA AGA TGG GCT CTG AGA AAT

Asp Gln Ile Glu Asp Asn Gly Ala Pro Ala Ala Arg Trp Ala Leu Arg Asn

GCC ATC ACA CGT TAC TTC ATG CAC GAC AGA CTC TGG CTG AAC GAT CCG GAC

Ala Ile Thr Arg Tyr Phe Met His Asp Arg Leu Trp Leu Asn Asp Pro Asp

TGC CTC ATC CTG AGA GAG GAA AAA ACA GAA CTG ACC CCA AAA GAG AGA GAG

Cys Leu Ile Leu Arg Glu Glu Lys Thr Glu Leu Thr Pro Lys Glu Arg Glu

CTC TAC TCG TAC ACC TGT GGG ATC CTC GAC AAC ATG ATC ATA GAA AGT GAC

Leu Tyr Ser Tyr Thr Cys Gly Ile Leu Asp Asn Met Ile Ile Glu Ser Asp

GAC CTG TCA CTT GTG AAA GAG CAC GGA AGG AAG GTT CTG AGA GAG ACA CTC

Asp Leu Ser Leu Val Lys Glu His Gly Arg Lys Val Leu Arg Glu Thr Leu

GAT CTT CTC GGG GGA AAG CCC CGT GTT CTG AAC ATC ATG ACA GAG GAT CTG

Asp Leu Leu Gly Gly Lys Pro Arg Val Leu Asn Ile Met Thr Glu Asp Leu

AAG TAC GAG ATC GTC TCG TCT GGC ACG ATC TCT GGA AAC ACC AGG CTC GTT

Lys Tyr Glu Ile Val Ser Ser Gly Thr Ile Ser Gly Asn Thr Arg Leu Val

GTC GAT CTC AAA AAC AGA GAG TAC CAT CTG GAA AAA GAG GGA AAG TCC TCT

Val Asp Leu Lys Asn Arg Glu Tyr His Leu Glu Lys Glu Gly Lys Ser Ser

CTG AGA AAG AAG GTT GTC AAA AGA GAA GAC GGA AGA AAC TTC TAC TTC TAC

Leu Arg Lys Lys Val Val Lys Arg Glu Asp Gly Arg Asn Phe Tyr Phe Tyr

GAA GAG GGT GAG AGA GAA TGA 1856

Glu Glu Gly Glu Arg Glu END

Thermotoga neapolitana (Clone # 56GP1) Glycosidase

1
ATG AGA AAA CTT GTG TTC TCA TTT TTG ATT GTG ACA TTG CCC ATC GTC CTC

Met Arg Lys Leu Val Phe Ser Phe Leu Ile Val Thr Leu Pro Ile Val Leu

TTT GCA AAC AGT GAT TTC GTG AAA GTG GAA AAC GGC AGG TTC ATA CTG AAC

Phe Ala Asn Ser Asp Phe Val Lys Val Glu Asn Gly Arg Phe Ile Leu Asn

GGA GAA GAG TTC AGA TTC GTT GGA AGC AAC AAC TAC TAC ATG CAC TAC AAG

Gly Glu Glu Phe Arg Phe Val Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys

AGC AAT CGA ATG ATA GAC AGT GTC CTT GAA AGT GCA AAA GCC ATG GGG GTG

Ser Asn Arg Met Ile Asp Ser Val Leu Glu Ser Ala Lys Ala Met Gly Val

AAG GTG CTC AGA ATT TGG GGA TTC CTC GAT GGT GAG AGT TAC TGC CGT GAC

Lys Val Leu Arg Ile Trp Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp

AAG AAC ACC TAC ATG CAC CCC GCA CCG GGA GTA TTT GGA TTG CCA GAG GGT

Lys Asn Thr Tyr Met His Pro Ala Pro Gly Val Phe Gly Leu Pro Glu Gly

ACG AAC GCT CAG GAC GGT TTT GAA AGA CTC GAC TAC ACG GTA GCG AAA GCA

Thr Asn Ala Gln Asp Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala

AAA GAA CTG GGC ATA AAG CTC ATA ATC GTT CTT GTG AAC AAC TGG GAC GAC

Lys Glu Leu Gly Ile Lys Leu Ile Ile Val Leu Val Asn Asn Trp Asp Asp

TTC GGT GGA ATG AAT CAA TAC GTG AGA TGG TTT GGG GGC ATC CAT CAC GAT

Phe Gly Gly Met Asn Gln Tyr Val Arg Trp Phe Gly Gly Ile His His Asp

GAC TTC TAC AGG AAC GAG AAG ATC AAA GAA GAA TAC AAA AAG TAC GTG TCT

Asp Phe Tyr Arg Asn Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser

TTC CTC ATA AAC AGG GTG AAC ACC TAC ACG GGT GTT CCT TAC AGG GAA GAG

Phe Leu Ile Asn Arg Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu

CCC ACC ATC ATG GCA TGG GAA CTG GCG AAC GAG CCC AGG TGT GAA ACG GAC

Pro Thr Ile Met Ala Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp

AAG TCT GGT AAC ACA CTC GTT GAA TGG GTA GAG GAG ATG AGT GCT TAC ATA

Lys Ser Gly Asn Thr Leu Val Glu Trp Val Glu Glu Met Ser Ala Tyr Ile

AAG AGT CTG GAT CCA AAC CAC CTG GTT GCC GTG GGA GAC GAG GGA TTC TTC

Lys Ser Leu Asp Pro Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe

AAC AAC TAC GAA GGC TTC AGA CCT TAC GGT GGA GAG GCT GAG TGG GCC TAC

Asn Asn Tyr Glu Gly Phe Arg Pro Tyr Gly Gly Glu Ala Glu Trp Ala Tyr

AAC GGA TGG TCC GGT GTT GAC TGG AAG AGA CTT CTG GAG ATA GAG ACG GTG

Asn Gly Trp Ser Gly Val Asp Trp Lys Arg Leu Leu Glu Ile Glu Thr Val

GAT TTT GGT ACG TTC CAT CTC TAC CCC TCC CAC TGG GGT GTG AGC CCT GAA

Asp Phe Gly Thr Phe His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu

AAC TAC GCA CAG TGG GGG GCA AAG TGG ATA GAA GAT CAC ATA AAG ATC GCA

Asn Tyr Ala Gln Trp Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala

AAA GAG GTT GGA AAA CCC GTC GTT CTG GAA GAG TAC GGT ATT CCC AAA AGT

Lys Glu Val Gly Lys Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser

GCC CCG GTC AAC AGG GTT GCC ATT TAC AAA TTG TGG AAC GAT CTG GTC TAC

Ala Pro Val Asn Arg Val Ala Ile Tyr Lys Leu Trp Asn Asp Leu Val Tyr

AAC CTC GGT GGA AAC GGT GCC ATG TTC TGG ATG CTC GCA GGA ATC GGT GAA

Asn Leu Gly Gly Asn Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu

GGA TGG GAC AGA GAC GAA AAG GGT TAC TAC CCC GAT TAC GAC GGC TTC AGA

Gly Trp Asp Arg Asp Glu Lys Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg

ATA GTG AAC GAT GAA AGT GAA GAG GCA AAG TTG ATC AGA GAG TAC GCG AAA

Ile Val Asn Asp Glu Ser Glu Glu Ala Lys Leu Ile Arg Glu Tyr Ala Lys

CTG TTC AGC ACG GGT GAG GAT ACG AGG GAA GAT ACC TGC ATG TTC ATC ACA

Leu Phe Ser Thr Gly Glu Asp Thr Arg Glu Asp Thr Cys Met Phe Ile Thr

CCA AAG GAT GGT CAG GAG ATC AAA AAG ACT GTG AAG GTG AGA GTG GGT GTC

Pro Lys Asp Gly Gln Glu Ile Lys Lys Thr Val Lys Val Arg Val Gly Val

TTC GAC TAC AGC AAC ACG TTC AAA GGA ATT TCC GTC GGG GTT GAA AAT CTG

Phe Asp Tyr Ser Asn Thr Phe Lys Gly Ile Ser Val Gly Val Glu Asn Leu

CTC TTT GAA GAT GAG ATA AAA CAT CTC GGA TAT GGA GTT TAC GGA TTC GAA

Leu Phe Glu Asp Glu Ile Lys His Leu Gly Tyr Gly Val Tyr Gly Phe Glu

TTT GAC ACA ACG CGG ATT TCA GAC GGA GAA CAC GAG ATG TTC CTT GAG GCA

Phe Asp Thr Thr Arg Ile Ser Asp Gly Glu His Glu Met Phe Leu Glu Ala

CAT TTC AGG GGA GAA ACG GTG AAA GAC ACA ATC AGG GTG AAA GTT GTG AAC

His Phe Arg Gly Glu Thr Val Lys Asp Thr Ile Arg Val Lys Val Val Asn

AGA GCG CAG TAT GTA CTC GCA GAA GAA GTG GAT TTT TCC AGA CCC GAA GAA

Arg Ala Gln Tyr Val Leu Ala Glu Glu Val Asp Phe Ser Arg Pro Glu Glu

GTC AAG AAC TGG TGG AAC AGC GGA ACA TGG CAG GCT GAG TTC AAA ACA CCC

Val Lys Asn Trp Trp Asn Ser Gly Thr Trp Gln Ala Glu Phe Lys Thr Pro

GAT ATA GAG TGG AAC GGT GAG GTG GGG AAC GGT GCT CTC CAG ATG AAC GTG

Asp Ile Glu Trp Asn Gly Glu Val Gly Asn Gly Ala Leu Gln Met Asn Val

GTG CTT CCC GGA AAG GGT GAC TGG GAA GAG GTG AGG GTG GTC AGG AAA TTC

Val Leu Pro Gly Lys Gly Asp Trp Glu Glu Val Arg Val Val Arg Lys Phe

GAT CAA CTC CCC GTG TGT GAG ATC CTC GAG TAC GAT ATC TAC ATA CCA GAC

Asp Gln Leu Pro Val Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asp

GTT GAA GGG CTT ACA GGA AGG CTC AGA CCG TAC GCG GTG CTG AAT CCC GGC

Val Glu Gly Leu Thr Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly

TGG GTG AAG ATA GGG CTC GAC ATG AAC AAC ACC TCG ATT GAC AGC GGA GAA

Trp Val Lys Ile Gly Leu Asp Met Asn Asn Thr Ser Ile Asp Ser Gly Glu

CTT GTC AGT TTC GAT GGC AAA AAG TAC AGA AAG TTC CAT GTG AGG ATC GAG

Leu Val Ser Phe Asp Gly Lys Lys Tyr Arg Lys Phe His Val Arg Ile Glu

TTC GAC AAG ACA CCT GGA GTG AAC GAG CTC CAC ATA GGT GTA GTT GGA GAC

Phe Asp Lys Thr Pro Gly Val Asn Glu Leu His Ile Gly Val Val Gly Asp

CAC CTG GAG TAT GAT GGG CCG ATT TTC ATC GAT AAT GTG AGG CTC TAT AAA

His Leu Glu Tyr Asp Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys

AAA TCT TCT TGA 2000

Lys Ser Ser END

ARP 2.3 (Alum Rock sulfur spring, Clone # 58GB3) Glycosidase

1

ATG CAT TTT AGC CCA CTA CAA TTG ATC CTC GTC TTA GTC ATT GTC ATT CTG

Met His Phe Ser Pro Leu Gln Leu Ile Leu Val Leu Val Ile Val Ile Leu

CTG TTT GGC ACC AAA AAA TTA CGC AAT ATG GGC GGC GAT TTA GGC GAA GCC

Leu Phe Gly Thr Lys Lys Leu Arg Asn Met Gly Gly Asp Leu Gly Glu Ala

TTC AAG AAT TTC AGA AAA GCA GTC AAA GAC GGC GAT GAT GCT GAA ACA CAA

Phe Lys Asn Phe Arg Lys Ala Val Lys Asp Gly Asp Asp Ala Glu Thr Gln

AAA GAT GTT GCT GTG CAA AAA GTT GAC CAA CAG CCA CCA GCA CAG CCC ATC

Lys Asp Val Ala Val Gln Lys Val Asp Gln Gln Pro Pro Ala Gln Pro Ile

254

CCA CAA GGT CGA GTC ATT GAT TCG GAA GCC AAG GAA AAG GAT AAG GTC TAA

Pro Gln Gly Arg Val Ile Asp Ser Glu Ala Lys Glu Lys Asp Lys Val END

AEPII 1a (Clone # 63GA3) Glycosidase

1
ATG GAA GGA CTT CGA GGA GGT GTG AGG ATG AAG TTC CCA TCT AAC TTT CTT

Met Glu Gly Leu Arg Gly Gly Val Arg Met Lys Phe Pro Ser Asn Phe Leu

TTT GGC TAC TCC TGG TCG GGC TTC CAG TTT GAA ATG GGT TTA CCT GGG AGT

Phe Gly Tyr Ser Trp Ser Gly Phe Gln Phe Glu Met Gly Leu Pro Gly Ser

GAA GTT GAG AGC GAC TGG TGG GCA TGG GTC CAC GAT AAG GAG AAC ATC TTC

Glu Val Glu Ser Asp Trp Trp Ala Trp Val His Asp Lys Glu Asn Ile Phe

TCG GGC CTA GTT AGC GGT GAC CTA CCA GAG AAC GGG CCT GCT TAC TGG CAC

Ser Gly Leu Val Ser Gly Asp Leu Pro Glu Asn Gly Pro Ala Tyr Trp His

CTC TAC AAG AAA GAC CAC GAC ATA GCT GAA AGC CTT GGC ATG GAC GCG ATA

Leu Tyr Lys Lys Asp His Asp Ile Ala Glu Ser Leu Gly Met Asp Ala Ile

AGA GGC GGA ATC GAG TGG GCG AGG ATC TTC CCA AAA CCC ACC TTT GAC GTG

Arg Gly Gly Ile Glu Trp Ala Arg Ile Phe Pro Lys Pro Thr Phe Asp Val

AAG GTT GAC GTG GAA AAG GAC GAA AAC GGG AAC ATA ATC TCC ATT GAC GTC

Lys Val Asp Val Glu Lys Asp Glu Asn Gly Asn Ile Ile Ser Ile Asp Val

CCG GAG AGC GCG ATA GAG GAG CTA GAA AAG CTT GCC AAC ATG GAT GCC CTC

Pro Glu Ser Ala Ile Glu Glu Leu Glu Lys Leu Ala Asn Met Asp Ala Leu

AAC CAC TAC CGC GAA ATC TAC TCG GAC TGG AAG GAG AGG GGC AAG ACC TTC

Asn His Tyr Arg Glu Ile Tyr Ser Asp Trp Lys Glu Arg Gly Lys Thr Phe

ATA TTG AAC CTC TAT CAC TGG CCC CTT CCC CTC TGG CTC CAC GAC CCG ATA

Ile Leu Asn Leu Tyr His Trp Pro Leu Pro Leu Trp Leu His Asp Pro Ile

GGC GTT AGA AAG CTC GGC CCT GAT AGA GCT CCC TCG GGC TGG CTG GAC GAG

Gly Val Arg Lys Leu Gly Pro Asp Arg Ala Pro Ser Gly Trp Leu Asp Glu

AGG AGC GTG GTG GAG TTC ACC AAG TTC GCT GCA TTC ATC GCC TAC CAC TTG

Arg Ser Val Val Glu Phe Thr Lys Phe Ala Ala Phe Ile Ala Tyr His Leu

GAT GAC CTC GTT GAC ATG TGG AGC ACG ATG AAC GAG CCG AAT GTG GTT TAC

Asp Asp Leu Val Asp Met Trp Ser Thr Met Asn Glu Pro Asn Val Val Tyr

GAG CAG GGT TAC ACG AGG CCT CAG TCG GGC TTT CCA CCG GGT TAT CTC AGC

Glu Gln Gly Tyr Thr Arg Pro Gln Ser Gly Phe Pro Pro Gly Tyr Leu Ser

CAC GAG GCC GCT GGA AAG GCG AAG CTC AAC CTC ATG CAG GCT CAC GCT AGA

His Glu Ala Ala Gly Lys Ala Lys Leu Asn Leu Met Gln Ala His Ala Arg

GCT TAC GAT GCG ATA AAA GAG CAC TCG GAC AAG CCC GTG GGG TTG ATA TAC
Ala Tyr Asp Ala Ile Lys Glu His Ser Asp Lys Pro Val Gly Leu Ile Tyr

TCC TTT GTC TGG CAC GAT GCC CTA AAC GAG GAA GCG GAG GAG ATT GTG AAG
Ser Phe Val Trp His Asp Ala Leu Asn Glu Glu Ala Glu Glu Ile Val Lys

GAG ATA AGG AGG AGA CAC TAC GAC TTC GTA ACC GGC CTT CAC TCC GGC TCA
Glu Ile Arg Arg Arg His Tyr Asp Phe Val Thr Gly Leu His Ser Gly Ser

TCG GAG TTC GGG GAG AGG GAG GAC TTC AAG GGG AAG ATC GAC TGG ATA GGC
Ser Glu Phe Gly Glu Arg Glu Asp Phe Lys Gly Lys Ile Asp Trp Ile Gly

GTG AAC TAC TAC ACT AGG GTT GCT TAC GAG ATG AGG AAC GGC CGC TTT ATG
Val Asn Tyr Tyr Thr Arg Val Ala Tyr Glu Met Arg Asn Gly Arg Phe Met

GCC CTA CCC GGG TAC GGC TAC ATG TGC GAG AGG AGT GGT TAC GCA AAA TCC
Ala Leu Pro Gly Tyr Gly Tyr Met Cys Glu Arg Ser Gly Tyr Ala Lys Ser

GGA AGG CCC GCG AGC GAT TTT GGC TGG GAG ACC TAT CCT GAG GGC CTC GAA
Gly Arg Pro Ala Ser Asp Phe Gly Trp Glu Thr Tyr Pro Glu Gly Leu Glu

AAC GTC CTG ATG GAT CTG AAG GAG CTC TAC GGC CTG CCA ATG ATG GTG ACG

```
Asn Val Leu Met Asp Leu Lys Glu Leu Tyr Gly Leu Pro Met Met Val Thr

GAG AAC GGG ATG GCG GAT ATG GCA GAC AGG CAC CGC TCT TAC TAC CTC GTG
Glu Asn Gly Met Ala Asp Met Ala Asp Arg His Arg Ser Tyr Tyr Leu Val

AGC CAC CTC GCG GCT ATC CAC AGG GCG ATG GAG AAG GGT GCC GAC GTT AGG
Ser His Leu Ala Ala Ile His Arg Ala Met Glu Lys Gly Ala Asp Val Arg

GGG TAC CTC CAC TGG TCT CTG ACC GAC AAC TAC GAG TGG GCG CAG GGC TTC
Gly Tyr Leu His Trp Ser Leu Thr Asp Asn Tyr Glu Trp Ala Gln Gly Phe

AGA ATG CGC TTT GGG CTG GTG ATG GTG GAC TTC GAG ACT AAG AAG CGC TAC
Arg Met Arg Phe Gly Leu Val Met Val Asp Phe Glu Thr Lys Lys Arg Tyr

ATA AGG CCG AGC GCA CTC GTC TTC AGG GAG ATA GCC ACG CAG AAG GAA ATA
Ile Arg Pro Ser Ala Leu Val Phe Arg Glu Ile Ala Thr Gln Lys Glu Ile

                                                                1478
CCC GAA GAG CTC TCC CAC CTA GCG AAC CTC GAA CTG GTA ACG AAG AAG TAA
Pro Glu Glu Leu Ser His Leu Ala Asn Leu Glu Leu Val Thr Lys Lys END
```

AEPII 1a (Clone # 63GA4) Glycosidase

1
ATG AAG TTC CCA TCT AAC TTT CTT TTT GGC TAC TCC TGG TCG GGC TTC CAG

Met Lys Phe Pro Ser Asn Phe Leu Phe Gly Tyr Ser Trp Ser Gly Phe Gln

TTT GAA ATG GGT TTA CCT GGG AGT GAA GTT GAG AGC GAC TGG TGG GCA TGG

Phe Glu Met Gly Leu Pro Gly Ser Glu Val Glu Ser Asp Trp Trp Ala Trp

GTC CAC GAT AAG GAG AAC ATC TTC TCG GGC CTA GTT AGC GGT GAC CTA CCA

Val His Asp Lys Glu Asn Ile Phe Ser Gly Leu Val Ser Gly Asp Leu Pro

GAG AAC GGG CCT GCT TAC TGG CAC CTC TAC AAG AAA GAC CAC GAC ATA GCT

Glu Asn Gly Pro Ala Tyr Trp His Leu Tyr Lys Lys Asp His Asp Ile Ala

GAA AGC CTT GGC ATG GAC GCG ATA AGA GGC GGA ATC GAG TGG GCG AGG ATC

Glu Ser Leu Gly Met Asp Ala Ile Arg Gly Gly Ile Glu Trp Ala Arg Ile

TTC CCA AAA CCC ACC TTT GAC GTG AAG GTT GAC GTG GAA AAG GAC GAA AAC

Phe Pro Lys Pro Thr Phe Asp Val Lys Val Asp Val Glu Lys Asp Glu Asn

GGG AAC ATA ATC TCC ATT GAC GTC CCG GAG AGC GCG ATA GAG GAG CTA GAA

Gly Asn Ile Ile Ser Ile Asp Val Pro Glu Ser Ala Ile Glu Glu Leu Glu

AAG CTT GCC AAC ATG GAT GCC CTC AAC CAC TAC CGC GAA ATC TAC TCG GAC
Lys Leu Ala Asn Met Asp Ala Leu Asn His Tyr Arg Glu Ile Tyr Ser Asp

TGG AAG GAG AGG GGC AAG ACC TTC ATA TTG AAC CTC TAT CAC TGG CCC CTT
Trp Lys Glu Arg Gly Lys Thr Phe Ile Leu Asn Leu Tyr His Trp Pro Leu

CCC CTC TGG CTC CAC GAC CCG ATA GGC GTT AGA AAG CTC GGC CCT GAT AGA
Pro Leu Trp Leu His Asp Pro Ile Gly Val Arg Lys Leu Gly Pro Asp Arg

GCT CCC TCG GGC TGG CTG GAC GAG AGG AGC GTG GTG GAG TTC ACC AAG TTC
Ala Pro Ser Gly Trp Leu Asp Glu Arg Ser Val Val Glu Phe Thr Lys Phe

GCT GCA TTC ATC GCC TAC CAC TTG GAT GAC CTC GTT GAC ATG TGG AGC ACG
Ala Ala Phe Ile Ala Tyr His Leu Asp Asp Leu Val Asp Met Trp Ser Thr

ATG AAC GAG CCG AAT GTG GTT TAC GAG CAG GGT TAC ACG AGG CCT CAG TCG
Met Asn Glu Pro Asn Val Val Tyr Glu Gln Gly Tyr Thr Arg Pro Gln Ser

GGC TTT CCA CCG GGT TAT CTC AGC CAC GAG GCC GCT GGA AAG GCG AAG CTC
Gly Phe Pro Pro Gly Tyr Leu Ser His Glu Ala Ala Gly Lys Ala Lys Leu

AAC CTC ATG CAG GCT CAC GCT AGA GCT TAC GAT GCG ATA AAA GAG CAC TCG
Asn Leu Met Gln Ala His Ala Arg Ala Tyr Asp Ala Ile Lys Glu His Ser

GAC AAG CCA GTT GGA GTT ATC TAC GCA TAT AAG TGG ATT GAT GCG GAG GAT

Asp Lys Pro Val Gly Val Ile Tyr Ala Tyr Lys Trp Ile Asp Ala Glu Asp

GAA GCT GCA GAG GAA TCC GTT CTG GAA CTC CGC AGG AGG GAT TAC GAC TTC

Glu Ala Ala Glu Glu Ser Val Leu Glu Leu Arg Arg Arg Asp Tyr Asp Phe

GTT GAT GGT CTC TAC TCA GGC AAG TCC CTG ACT GCA GGT GAG AGG GAG GAC

Val Asp Gly Leu Tyr Ser Gly Lys Ser Leu Thr Ala Gly Glu Arg Glu Asp

TTC AAA GGC AGG GTC GAC TGG GTT GGC GTC AAC TAC TAC TCC CGC CTG CTC

Phe Lys Gly Arg Val Asp Trp Val Gly Val Asn Tyr Tyr Ser Arg Leu Leu

TTT GGA AAG GCC GGA GAT TCA GTG AGA TTA CTT GAG GGC TAC GGT TTT GTC

Phe Gly Lys Ala Gly Asp Ser Val Arg Leu Leu Glu Gly Tyr Gly Phe Val

TCC CCG AGG GGT GGC TAC GCC AAA TCG GGA AGG CCT GCG AGC GAT TTT GGC

Ser Pro Arg Gly Gly Tyr Ala Lys Ser Gly Arg Pro Ala Ser Asp Phe Gly

TGG GAG ATT TAT CCT GAG GGC CTC GAA AAG CTC CTG GTT GAG CTG AGT GGC

Trp Glu Ile Tyr Pro Glu Gly Leu Glu Lys Leu Leu Val Glu Leu Ser Gly

AGG TAC GAG CTT CCG CTC TTC ATA ACG GAG AAT GGT ATG GCT GAT GCT GTC

Arg Tyr Glu Leu Pro Leu Phe Ile Thr Glu Asn Gly Met Ala Asp Ala Val

GAT AGG TAC AGG CCT TAC TAC CTC GTG AGC CAC CTC GCG GCT ATC CAC AGG

Asp Arg Tyr Arg Pro Tyr Tyr Leu Val Ser His Leu Ala Ala Ile His Arg

GCG ATG GAG AAG GGT GCC GAC ATT AGG GGG TAC CTC CAC TGG TCT CTG ACC

Ala Met Glu Lys Gly Ala Asp Ile Arg Gly Tyr Leu His Trp Ser Leu Thr

GAC AAC TAC GAG TGG GCG CAG GGC TTC AGA ATG CGC TTT GGG CTG GTG ATG

Asp Asn Tyr Glu Trp Ala Gln Gly Phe Arg Met Arg Phe Gly Leu Val Met

GTG GAC TTC GAG ACT AAG AAG CGC TAC TTG AGG CCG AGC GCA CTC GTC TTC

Val Asp Phe Glu Thr Lys Lys Arg Tyr Leu Arg Pro Ser Ala Leu Val Phe

AGG GAA ATA GCC ACG CGG AAG GAA ATA CCC GAA GAG CTT GAA CAC CTT GCC

Arg Glu Ile Ala Thr Arg Lys Glu Ile Pro Glu Glu Leu Glu His Leu Ala

GAT GTG GAT GCA ATC ATT GCT CGG TGA 1454

Asp Val Asp Ala Ile Ile Ala Arg END

AEPII 1a (Clone # 63GA9) Glycosidase

1

ATG CTA CCA GAA GAG TTC CTA TGG GGC GTT GGG CAG TCA GGC TTT CAG TTC

Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe

GAA ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG

Glu Met Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp

AAG TGG GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC

Lys Trp Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp

CTT CCC GAG GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG

Leu Pro Glu Asp Gly Ile Asn Asn Tyr Glu Leu Phe Glu Asn Asp His Lys

CTC GCT AAA GGC CTT GGA CTC AAC GCA TAC AGG ATT GGA ATA GAG TGG AGC

Leu Ala Lys Gly Leu Gly Leu Asn Ala Tyr Arg Ile Gly Ile Glu Trp Ser

AGA ATC TTT CCC TGG CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC

Arg Ile Phe Pro Trp Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp

ACT TAC GGT TTA GTA AAG GAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA

Thr Tyr Gly Leu Val Lys Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu

CTC GAC AGG CTG GCC AAC AAG GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT
Leu Asp Arg Leu Ala Asn Lys Glu Glu Val Met Tyr Tyr Arg Arg Val Ile

CAG CAT TTG AGG GAG CTC GGC TTC AAG GTC TTC GTT AAC CTC AAC CAC TTC
Gln His Leu Arg Glu Leu Gly Phe Lys Val Phe Val Asn Leu Asn His Phe

ACG CTT CCA ATA TGG CTC CAC GAC CCG ATA GTG GCA AGG GAG AAG GCC CTC
Thr Leu Pro Ile Trp Leu His Asp Pro Ile Val Ala Arg Glu Lys Ala Leu

ACA AAC GAC AGA ATC GGC TGG GTC TCC CAG AGG ACA GTT GTT GAG TTT GCC
Thr Asn Asp Arg Ile Gly Trp Val Ser Gln Arg Thr Val Val Glu Phe Ala

AAG TAT GCT GCT TAC ATC GCC CAT GCG CTC GGA GAC CTC GTG GAC ACA TGG
Lys Tyr Ala Ala Tyr Ile Ala His Ala Leu Gly Asp Leu Val Asp Thr Trp

AGC ACC TTC AAC GAA CCT ATG GTA GTT GTG GAG CTC GGC TAC CTC GCC CCC
Ser Thr Phe Asn Glu Pro Met Val Val Val Glu Leu Gly Tyr Leu Ala Pro

TAC TCA GGA TTT CCC CCG GGA GTC ATG AAC CCC GAG GCC GCG AAG CTG GCG
Tyr Ser Gly Phe Pro Pro Gly Val Met Asn Pro Glu Ala Ala Lys Leu Ala

ATC CTC AAC ATG ATA AAC GCC CAC GCC TTG GCA TAT AAG ATG ATA AAG AGG
Ile Leu Asn Met Ile Asn Ala His Ala Leu Ala Tyr Lys Met Ile Lys Arg

TTC GAC ACC AAG AAG GCC GAT GAG GAT AGC AAG TCC CCT GCG GAC GTT GGC

Phe Asp Thr Lys Lys Ala Asp Glu Asp Ser Lys Ser Pro Ala Asp Val Gly

ATA ATC TAC AAC AAC ATC GGT GTT GCC TAC CCT AAA GAC CCT AAC GAT CCC

Ile Ile Tyr Asn Asn Ile Gly Val Ala Tyr Pro Lys Asp Pro Asn Asp Pro

AAG GAC GTT AAA GCA GCC GAA AAC GAC AAC TAC TTC CAC AGC GGA CTG TTC

Lys Asp Val Lys Ala Ala Glu Asn Asp Asn Tyr Phe His Ser Gly Leu Phe

TTT GAT GCC ATC CAC AAG GGT AAG CTC AAC ATA GAG TTC GAC GGC GAA AAC

Phe Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe Asp Gly Glu Asn

TTT GTA AAA GTT AGA CAC CTA AAA GGC AAT GAC TGG ATA GGC CTC AAC TAC

Phe Val Lys Val Arg His Leu Lys Gly Asn Asp Trp Ile Gly Leu Asn Tyr

TAC ACC CGC GAG GTT GTT AGA TAT TCG GAG CCC AAG TTC CCA AGT ATA CCC

Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro

CTC ATA TCC TTC AAG GGC GTT CCC AAC TAC GGC TAC TCC TGC AGG CCC GGC

Leu Ile Ser Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Gly

ACG ACC TCC GCC GAT GGC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT

Thr Thr Ser Ala Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr

CCC CAG GGA ATC TAC GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT
Pro Gln Gly Ile Tyr Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro

GTT TAC GTC ACC GAG AAC GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA
Val Tyr Val Thr Glu Asn Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro

TAC TAC ATA GTC AGC CAC GTC TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA
Tyr Tyr Ile Val Ser His Val Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly

TAC CCC GTA AAA GGC TAC ATG TAC TGG GCG CTT ACG GAT AAC TAC GAG TGG
Tyr Pro Val Lys Gly Tyr Met Tyr Trp Ala Leu Thr Asp Asn Tyr Glu Trp

GCC CTC GGC TTC AGC ATG AGG TTT GGT CTC TAC AAG GTC GAC CTC ATC TCC
Ala Leu Gly Phe Ser Met Arg Phe Gly Leu Tyr Lys Val Asp Leu Ile Ser

AAG GAG AGG ATC CCG AGG GAG AGA AGC GTT GAG ATA TAT CGC AGG ATA GTG
Lys Glu Arg Ile Pro Arg Glu Arg Ser Val Glu Ile Tyr Arg Arg Ile Val

CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG GAG TTC CTG AAG GGT GAG
Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu Glu Phe Leu Lys Gly Glu

GAG AAA TGA 1538
Glu Lys END

AEPII 1a (Clone # 63GB1) Glycosidase

1
ATG CTA CCA GAA GAG TTC CTA TGG GGC GTT GGG CAG TCA GGC TTT CAG TTC
Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe

GAA ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG
Glu Met Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp

AAG TGG GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC
Lys Trp Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp

CTT CCC GAG GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG
Leu Pro Glu Asp Gly Ile Asn Asn Tyr Glu Leu Phe Glu Asn Asp His Lys

CTC GCT AAA GGC CTT GGA CTC AAC GCA TAC GGG ATT GGA ATA GAG TGG AGC
Leu Ala Lys Gly Leu Gly Leu Asn Ala Tyr Gly Ile Gly Ile Glu Trp Ser

AGA ATC TTT CCC TGG CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC
Arg Ile Phe Pro Trp Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp

ACT TAC GGT TTA GTA AAG GAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA
Thr Tyr Gly Leu Val Lys Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu

CTC GAC AGG CTG GCC AAC AAG GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT

Leu Asp Arg Leu Ala Asn Lys Glu Glu Val Met Tyr Tyr Arg Arg Val Ile

CAG CAT TTG AGG GAG CTC GGC TTC AAG GTC TTC GTT AAC CTC AAC CAC TTC

Gln His Leu Arg Glu Leu Gly Phe Lys Val Phe Val Asn Leu Asn His Phe

ACG CTT CCA ATA TGG CTC CAC GAC CCG ATA GTG GCA AGG GAG AAG GCC CTC

Thr Leu Pro Ile Trp Leu His Asp Pro Ile Val Ala Arg Glu Lys Ala Leu

ACA AAC GAC AGA ATC GGC TGG GTC TCC CAG AGG ACA GTT GTT GAG TTT GCC

Thr Asn Asp Arg Ile Gly Trp Val Ser Gln Arg Thr Val Val Glu Phe Ala

AAG TAT GCT GCT TAC ATC GCC CAT GCG CTC GGA GAC CTC GTG GAC ACA TGG

Lys Tyr Ala Ala Tyr Ile Ala His Ala Leu Gly Asp Leu Val Asp Thr Trp

AGC ACC TTC AAC GAA CCT ATG GTA GTT GTG GAG CTC GGA TAC CTC GCC CCC

Ser Thr Phe Asn Glu Pro Met Val Val Val Glu Leu Gly Tyr Leu Ala Pro

TAC TCA GGA TTT CCC CCG GGA GTC ATG AAC CCC GAG GCC GCG AAG CTG GCG

Tyr Ser Gly Phe Pro Pro Gly Val Met Asn Pro Glu Ala Ala Lys Leu Ala

ATC CTC AAC ATG ATA AAC GCC CAC GCC TTG GCA TAT AAG ATG ATA AAG AGG

Ile Leu Asn Met Ile Asn Ala His Ala Leu Ala Tyr Lys Met Ile Lys Arg

TTC GAC ACC AAG AAG GCC GAT GAG GAT AGC AAG TCC CCT GCG GAC GTT GGC

Phe Asp Thr Lys Lys Ala Asp Glu Asp Ser Lys Ser Pro Ala Asp Val Gly

ATA ATC TAC AAC AAC ATC GGT GTT GCC TAC CCT AAA GAC CCT AAC GAT CCC

Ile Ile Tyr Asn Asn Ile Gly Val Ala Tyr Pro Lys Asp Pro Asn Asp Pro

AAG GAC GTT AAA GCA GCC GAA AAC GAC AAC TAC TTC CAC AGC GGA CTG TTC

Lys Asp Val Lys Ala Ala Glu Asn Asp Asn Tyr Phe His Ser Gly Leu Phe

TTT GAT GCC ATC CAC AAG GGT AAG CTC AAC ATA GAG TTC GAC GGC GAA AAC

Phe Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe Asp Gly Glu Asn

TTT GTA AAA GTT AGA CAC CTA AAA GGC AAT GAC TGG ATA GGC CTC AAC TAC

Phe Val Lys Val Arg His Leu Lys Gly Asn Asp Trp Ile Gly Leu Asn Tyr

TAC ACC CGC GAG GTT GTT AGA TAT TCG GAG CCC AAG TTC CCA AGT ATA CCC

Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro

CTC ATA TCC TTC AAG GGC GTT CCC AAC TAC GGC TAC TCC TGC AGG CCC GGC

Leu Ile Ser Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Gly

ACG ACC TCC GCC GAT GGC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT

Thr Thr Ser Ala Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr

CCC CAG GGA ATC TAC GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT
Pro Gln Gly Ile Tyr Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro

GTT TAC GTC ACC GAG AAC GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA
Val Tyr Val Thr Glu Asn Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro

TAC TAC ATA GTC AGC CAC GTC TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA
Tyr Tyr Ile Val Ser His Val Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly

TAC CCC GTA AAA GGC TAC ATG TAC TGG GCG CTT ACG GAT AAC TAC GAG TGG
Tyr Pro Val Lys Gly Tyr Met Tyr Trp Ala Leu Thr Asp Asn Tyr Glu Trp

GCC CTC GGC TTC AGC ATG AGG TTT GGT CTC TAC AAG GTC GAC CTC ATC TCC
Ala Leu Gly Phe Ser Met Arg Phe Gly Leu Tyr Lys Val Asp Leu Ile Ser

AAG GAG AGG ATC CCG AGG GAG AGA AGC GTT GAG ATA TAT CGC AGG ATA GTG
Lys Glu Arg Ile Pro Arg Glu Arg Ser Val Glu Ile Tyr Arg Arg Ile Val

CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG GAG TTC CTG AAG GGT GAG
Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu Glu Phe Leu Lys Gly Glu

```
GAG AAA TGA        1538
Glu Lys END
```

AEPII 1a (Clone # 63GP1) Glycosidase

1
ATG CGT CCA TTC TTG TTA ATT TCT ATT TTG GAC TTT CGA GTT GCT GAC TAC
Met Arg Pro Phe Leu Leu Ile Ser Ile Leu Asp Phe Arg Val Ala Asp Tyr

CTC CAA CGT AAC ATA AAG ACA CAA AAC CAA TAT TGG GCA TTG TGC GTA GTA
Leu Gln Arg Asn Ile Lys Thr Gln Asn Gln Tyr Trp Ala Leu Cys Val Val

ATG TTC TCC AAT GTT CTT AGA TGG CAA AAC TTA AAT ATT TCA CCA GCG GTG
Met Phe Ser Asn Val Leu Arg Trp Gln Asn Leu Asn Ile Ser Pro Ala Val

ATA CAT AGA GAC ACC GCT GAA CAC AGA GGT GAT TCC ATG AAG AAG TTT GTC
Ile His Arg Asp Thr Ala Glu His Arg Gly Asp Ser Met Lys Lys Phe Val

GCC CTG TTC ATA ACC ATG TTT TTC GTA GTG AGC ATG GCA GTC GTT GCA CAG
Ala Leu Phe Ile Thr Met Phe Phe Val Val Ser Met Ala Val Val Ala Gln

CCA GCT AGC GCC GCA AAG TAT TCC GAG CTC GAA GAA GGC GGC GTT ATA ATG
Pro Ala Ser Ala Ala Lys Tyr Ser Glu Leu Glu Glu Gly Gly Val Ile Met

CAG GCC TTC TAC TGG GAC GTC CCA GGT GGA GGA ATC TGG TGG GAC ACC ATC
Gln Ala Phe Tyr Trp Asp Val Pro Gly Gly Gly Ile Trp Trp Asp Thr Ile

AGG AGC AAG ATA CCG GAG TGG TAC GAG GCG GGA ATA TCC GCC ATT TGG ATT

Arg Ser Lys Ile Pro Glu Trp Tyr Glu Ala Gly Ile Ser Ala Ile Trp Ile

CCG CCA GCC AGC AAG GGG ATG AGC GGC GGT TAC TCG ATG GGC TAC GAT CCC

Pro Pro Ala Ser Lys Gly Met Ser Gly Gly Tyr Ser Met Gly Tyr Asp Pro

TAC GAT TTC TTT GAC CTC GGC GAG TAC AAC CAG AAG GGA ACC ATC GAA ACG

Tyr Asp Phe Phe Asp Leu Gly Glu Tyr Asn Gln Lys Gly Thr Ile Glu Thr

CGC TTT GGC TCT AAA CAG GAG CTC ATC AAT ATG ATA AAC ACG GCC CAT GCC

Arg Phe Gly Ser Lys Gln Glu Leu Ile Asn Met Ile Asn Thr Ala His Ala

TAC GGC ATA AAG GTC ATA GCG GAC ATC GTC ATA AAC CAC CGC GCA GGC GGA

Tyr Gly Ile Lys Val Ile Ala Asp Ile Val Ile Asn His Arg Ala Gly Gly

GAC CTC GAG TGG AAC CCG TTC GTT GGG GAC TAC ACC TGG ACG GAC TTC TCA

Asp Leu Glu Trp Asn Pro Phe Val Gly Asp Tyr Thr Trp Thr Asp Phe Ser

AAG GTG GCC TCG GGC AAA TAT ACT GCC AAC TAC CTC GAC TTC CAC CCC AAC

Lys Val Ala Ser Gly Lys Tyr Thr Ala Asn Tyr Leu Asp Phe His Pro Asn

GAG GTC AAG TGC TGT GAC GAG GGC ACA TTT GGA GGC TTC CCA GAC ATA GCC

Glu Val Lys Cys Cys Asp Glu Gly Thr Phe Gly Gly Phe Pro Asp Ile Ala

CAC GAG AAG AGC TGG GAC CAG CAC TGG CTC TGG GCG AGC GAT GAG AGC TAC

His Glu Lys Ser Trp Asp Gln His Trp Leu Trp Ala Ser Asp Glu Ser Tyr

GCC GCC TAC CTA AGG AGC ATC GGC GTT GAT GCC TGG CGC TTT GAC TAC GTG

Ala Ala Tyr Leu Arg Ser Ile Gly Val Asp Ala Trp Arg Phe Asp Tyr Val

AAG GGC TAC GGA GCG TGG GTC GTC AAG GAC TGG CTC AAC TGG TGG GGC GGC

Lys Gly Tyr Gly Ala Trp Val Val Lys Asp Trp Leu Asn Trp Trp Gly Gly

TGG GCC GTT GGC GAG TAC TGG GAC ACC AAC GTT GAT GCA CTC CTC AAC TGG

Trp Ala Val Gly Glu Tyr Trp Asp Thr Asn Val Asp Ala Leu Leu Asn Trp

GCC TAC TCG AGC GGC GCC AAG GTC TTC GAC TTC CCG CTC TAC TAC AAG ATG

Ala Tyr Ser Ser Gly Ala Lys Val Phe Asp Phe Pro Leu Tyr Tyr Lys Met

GAT GAG GCC TTT GAC AAC AAA AAC ATT CCA GCG CTC GTC TCT GCC CTT CAG

Asp Glu Ala Phe Asp Asn Lys Asn Ile Pro Ala Leu Val Ser Ala Leu Gln

AAC GGC CAG ACT GTT GTC TCC CGC GAC CCG TTC AAG GCC GTA ACC TTT GTA

Asn Gly Gln Thr Val Val Ser Arg Asp Pro Phe Lys Ala Val Thr Phe Val

GCA AAC CAC GAC ACC GAT ATA ATC TGG AAC AAG TAC CTT GCT TAT GCT TTC

Ala Asn His Asp Thr Asp Ile Ile Trp Asn Lys Tyr Leu Ala Tyr Ala Phe

ATC CTC ACC TAC GAA GGC CAG CCC GTC ATA TTT TAC CGC GAC TAC GAG GAG
Ile Leu Thr Tyr Glu Gly Gln Pro Val Ile Phe Tyr Arg Asp Tyr Glu Glu

TGG CTC AAC AAG GAC AGG TTG AAC AAC CTC ATA TGG ATA CAC GAC CAC CTC
Trp Leu Asn Lys Asp Arg Leu Asn Asn Leu Ile Trp Ile His Asp His Leu

GCA GGT GGA AGC ACG AGC ATA GTT TAC TAC GAC AGC GAC GAG ATG ATT TTC
Ala Gly Gly Ser Thr Ser Ile Val Tyr Tyr Asp Ser Asp Glu Met Ile Phe

GTG AGG AAC GGC TAT GGA AGC AAG CCT GGC CTT ATA ACT TAC ATC AAC CTC
Val Arg Asn Gly Tyr Gly Ser Lys Pro Gly Leu Ile Thr Tyr Ile Asn Leu

GGC TCG AGC AAG GTT GGA AGG TGG GTT TAT GTG CCG AAG TTC GCG GGC GCG
Gly Ser Ser Lys Val Gly Arg Trp Val Tyr Val Pro Lys Phe Ala Gly Ala

TGC ATC CAC GAG TAT ACT GGT AAC CTC GGA GGC TGG GTA GAC AAG TAC GTC
Cys Ile His Glu Tyr Thr Gly Asn Leu Gly Gly Trp Val Asp Lys Tyr Val

TAC TCA AGC GGC TGG GTC TAT TTC GAA GCT CCA GCT TAC GAC CCT GCC AAC
Tyr Ser Ser Gly Trp Val Tyr Phe Glu Ala Pro Ala Tyr Asp Pro Ala Asn

GGG CAG TAT GGC TAC TCC GTG TGG AGC TAT TGC GGT GTT GGG TGA 1574

Gly Gln Tyr Gly Tyr Ser Val Trp Ser Tyr Cys Gly Val Gly END

AEPII 1a (Clone # 63GP2) Glycosidase

1
ATG ATA AAC GTT GCA ACG GGA GAG GAG ACC CCA ATA CAC CTC TTT GGA GTC
Met Ile Asn Val Ala Thr Gly Glu Glu Thr Pro Ile His Leu Phe Gly Val

AAC TGG TTC GGC TTT GAG ACA CCG AAC TAC GTT GTT CAC GGC CTA TGG AGT
Asn Trp Phe Gly Phe Glu Thr Pro Asn Tyr Val Val His Gly Leu Trp Ser

AGG AAC TGG GAG GAC ATG CTC CTC CAG ATC AAG AGC CTT GGC TTC AAT GCG
Arg Asn Trp Glu Asp Met Leu Leu Gln Ile Lys Ser Leu Gly Phe Asn Ala

ATA AGG CTT CCC TTC TGT ACC CAG TCA GTA AAA CCG GGG ACG ATG CCA ACG
Ile Arg Leu Pro Phe Cys Thr Gln Ser Val Lys Pro Gly Thr Met Pro Thr

GCG ATT GAC TAC GCC AAG AAC CCA GAC CTC CAG GGT CTT GAC AGC GTC CAG
Ala Ile Asp Tyr Ala Lys Asn Pro Asp Leu Gln Gly Leu Asp Ser Val Gln

ATA ATG GAG AAA ATA ATC AAG AAG GCT GGA GAC CTG GGC ATA TTC GTG CTC
Ile Met Glu Lys Ile Ile Lys Lys Ala Gly Asp Leu Gly Ile Phe Val Leu

CTC GAC TAC CAC AGA ATA GGA TGC AAC TTC ATA GAA CCC CTA TGG TAC ACC
Leu Asp Tyr His Arg Ile Gly Cys Asn Phe Ile Glu Pro Leu Trp Tyr Thr

GAC AGC TTC TCG GAG CAG GAC TAC ATA AAC ACC TGG GTT GAA GTC GCC CAG

Asp Ser Phe Ser Glu Gln Asp Tyr Ile Asn Thr Trp Val Glu Val Ala Gln

AGG TTC GGC AAG TAC TGG AAC GTT ATC GGC GCG GAC CTG AAG AAC GAA CCC

Arg Phe Gly Lys Tyr Trp Asn Val Ile Gly Ala Asp Leu Lys Asn Glu Pro

CAC AGC TCA AGC CCC GCA CCT GCC GCC TAC ACT GAC GGA AGT GGG GCC ACG

His Ser Ser Ser Pro Ala Pro Ala Ala Tyr Thr Asp Gly Ser Gly Ala Thr

TGG GGA ATG GGC AAC AAC GCC ACC GAC TGG AAC CTG GCG GCT GAG AGG ATA

Trp Gly Met Gly Asn Asn Ala Thr Asp Trp Asn Leu Ala Ala Glu Arg Ile

GGA AGG GCA ATT CTG GAG GTT GCC CCA CAA TGG GTT ATA TTT GTT GAG GGA

Gly Arg Ala Ile Leu Glu Val Ala Pro Gln Trp Val Ile Phe Val Glu Gly

ACC CAG TTC ACC ACC CCC GAG ATA GAC GGT AGG TAC AAG TGG GGC CAC AAC

Thr Gln Phe Thr Thr Pro Glu Ile Asp Gly Arg Tyr Lys Trp Gly His Asn

GCC TGG TGG GGC GGA AAC CTT ATG GGT GTT AGG AAG TAC CCA GTT AAC CTG

Ala Trp Trp Gly Gly Asn Leu Met Gly Val Arg Lys Tyr Pro Val Asn Leu

CCC AGG GAC AAG CTT GTT TAC AGC CCC CAA GTT TAC GGT CCA GAC GTT TAC

Pro Arg Asp Lys Leu Val Tyr Ser Pro Gln Val Tyr Gly Pro Asp Val Tyr

GAC CAG CCC TAC TTT GAC CCC GGT GAG GGG TTC CCC GAC AAC CTC CCC GAA

Asp Gln Pro Tyr Phe Asp Pro Gly Glu Gly Phe Pro Asp Asn Leu Pro Glu

ATA TGG TAC CAC CAC TTC GGC TAC GTA AAG CTT GAT CTC GGT TAC CCT GTT

Ile Trp Tyr His His Phe Gly Tyr Val Lys Leu Asp Leu Gly Tyr Pro Val

GTT ATA GGT GAG TTC GGA GGC AAG TAC GGC CAT GGG GGA GAC CCG AGG GAT

Val Ile Gly Glu Phe Gly Gly Lys Tyr Gly His Gly Gly Asp Pro Arg Asp

GTC ACT TGG CAG AAC AAG ATA ATA GAC TGG ATG ATC CAG AAC AAA TTC TGT

Val Thr Trp Gln Asn Lys Ile Ile Asp Trp Met Ile Gln Asn Lys Phe Cys

GAC TTC TTC TAC TGG AGC TGG AAC CCA AAC AGC GGT GAC ACC GGT GGA ATT

Asp Phe Phe Tyr Trp Ser Trp Asn Pro Asn Ser Gly Asp Thr Gly Gly Ile

CTG AAG GAT GAC TGG ACG ACA ATA TGG GAG GAC AAG TAC AAC AAC CTG AAG

Leu Lys Asp Asp Trp Thr Thr Ile Trp Glu Asp Lys Tyr Asn Asn Leu Lys

AGG CTC ATG GAC AGC TGT TCT GGA AAC GCC ACT GCC CCG TCC GTC CCC ACG

Arg Leu Met Asp Ser Cys Ser Gly Asn Ala Thr Ala Pro Ser Val Pro Thr

ACA ACT ACA ACA ACA AGC ACA CCG CCA ACG ACC ACA ACG ACT ACA ACA TCC

Thr Thr Thr Thr Thr Ser Thr Pro Pro Thr Thr Thr Thr Thr Thr Ser

ACT CCA ACG ACC ACT ACC CAG ACC CCG ACC ACC ACT ACT CCA ACT ACG ACA

Thr Pro Thr Thr Thr Thr Gln Thr Pro Thr Thr Thr Thr Pro Thr Thr Thr

ACC ACC ACG ACC ACA ACT CCT TCA AAT AAC GTC CCA TTT GAA ATT GTG AAC

Thr Thr Thr Thr Thr Thr Pro Ser Asn Asn Val Pro Phe Glu Ile Val Asn

GTT CTC CCG ACT AGC TCC CAG TAC GAG GGA ACC AGC GTG GAG GTT GTA TGT

Val Leu Pro Thr Ser Ser Gln Tyr Glu Gly Thr Ser Val Glu Val Val Cys

GAT GGA ACC CAG TGT GCC TCC AGC GTT TGG GGA GCT CCG AAC CTC TGG GGA

Asp Gly Thr Gln Cys Ala Ser Ser Val Trp Gly Ala Pro Asn Leu Trp Gly

GTC GTT AAA ATC GGA AAC GCC ACC ATG GAC CCC AAC GTT TGG GGC TGG GAG

Val Val Lys Ile Gly Asn Ala Thr Met Asp Pro Asn Val Trp Gly Trp Glu

GAC GTT TAC AAG ACT GCA CCC CAG GAC ATT GGA ACC GGC AGC ACA AAG ATG

Asp Val Tyr Lys Thr Ala Pro Gln Asp Ile Gly Thr Gly Ser Thr Lys Met

GAG ATA AGG AAC GGG GTG CTC AAG GTT ACA AAC CTC TGG AAC ATC AAC ATG

Glu Ile Arg Asn Gly Val Leu Lys Val Thr Asn Leu Trp Asn Ile Asn Met

AEPII 1a (Clone # 63GP4) Glycosidase

1

GCT GGA GTG GGT GAG CAA CGG GAT AAC CTA CCA GAT ATT CCC CGA CAG GTT

Ala Gly Val Gly Glu Gln Arg Asp Asn Leu Pro Asp Ile Pro Arg Gln Val

CAA CAA CGG AAA CAG GAG CAA CGA TGC CCT AGC TTT GGA CCA CGA CGA GCT

Gln Gln Arg Lys Gln Glu Gln Arg Cys Pro Ser Phe Gly Pro Arg Arg Ala

AAT TCT GAA CCA GGT CAA TCC AGG CAA ACC AAT CCT CTC CAA CTG GAG CGA

Asn Ser Glu Pro Gly Gln Ser Arg Gln Thr Asn Pro Leu Gln Leu Glu Arg

CCC TAT AAC GCC CCT CCA CTG CTG CCA CCA GTA CTT CGG CGG CGA CAT AAA

Pro Tyr Asn Ala Pro Pro Leu Leu Pro Pro Val Leu Arg Arg Arg His Lys

GGG AAT AAC GGA GAA GCT CGA CTA CCT TCA GAG CCT AGG TGT TAC TAT AAT

Gly Asn Asn Gly Glu Ala Arg Leu Pro Ser Glu Pro Arg Cys Tyr Tyr Asn

CTA CCT CAA CCC GAT TTT CCT CTC GGG AAG CGC CCA CGG CTA CGA CAC CTA

Leu Pro Gln Pro Asp Phe Pro Leu Gly Lys Arg Pro Arg Leu Arg His Leu

CGA CTA CTA CCG GCT TGA CCC CAA GTT CGG GAC CGA GGA GGA GCT GAG AGA

Arg Leu Leu Pro Ala End Pro Gln Val Arg Asp Arg Gly Gly Ala Glu Arg

CAT CCG AAG TAT AAC ACA ATG GCA TAC CCG GAG GTC ATA TAC GGC GCC AAG

His Pro Lys Tyr Asn Thr Met Ala Tyr Pro Glu Val Ile Tyr Gly Ala Lys

CCT TGG GGC AAC CAG CCA ATA AAC GCT CCG AAC TTC GTG CTC CCG ATA AAG

Pro Trp Gly Asn Gln Pro Ile Asn Ala Pro Asn Phe Val Leu Pro Ile Lys

GTC TCC CAG CTT CCG AGG ATA CTC GTT GAC ACA AAG TAC ACG CTC GAA AAG

Val Ser Gln Leu Pro Arg Ile Leu Val Asp Thr Lys Tyr Thr Leu Glu Lys

AGC TTC CCG GGA AAC AAC TTC GCC TTT GAG GCC TGG CTC TTC AAG GAT GCC

Ser Phe Pro Gly Asn Asn Phe Ala Phe Glu Ala Trp Leu Phe Lys Asp Ala

AAC AAC ATG AGG GCA CCA GGC CAG GGG GAC TAC GAG AGG AAT TCC GCC GAT

Asn Asn Met Arg Ala Pro Gly Gln Gly Asp Tyr Glu Arg Asn Ser Ala Asp

ACT GAC GGG CTC CAG GAG TCG TCG CCA CCA ATC CCC ATA TGG AAA CCG TCG

Thr Asp Gly Leu Gln Glu Ser Ser Pro Pro Ile Pro Ile Trp Lys Pro Ser

1886

ATA AGC TTG CGG CCG CCA CCG CGG TGG AGC TCC AGC TTT TGT TCC CTT TAA

Ile Ser Leu Arg Pro Pro Pro Arg Trp Ser Ser Ser Phe Cys Ser Leu END

GTT CCT CGA TGA GGC ACA CAG GCG GGG AAT GAG GGT AAT TTT CGA TTT TGT

Val Pro Arg End Gly Thr Gln Ala Gly Asn Glu Gly Asn Phe Arg Phe Cys

GCC CAA CCA CTG CGG CAT AGG GAA TCC AGC CTT CCT AGA AGT TTG GAA GAA

Ala Gln Pro Leu Arg His Arg Glu Ser Ser Leu Pro Arg Ser Leu Glu Glu

GGG CAA CGA AAG CCC ATA CTG GGA CTG GTT CTT CGT CAA GAA GTG GCC GTT

Gly Gln Arg Lys Pro Ile Leu Gly Leu Val Leu Arg Gln Glu Val Ala Val

CAA GCT CGG CGA TGG GAA CGC CTA CGT CGG CTG GTG GGG CTT TGG GAG CCT

Gln Ala Arg Arg Trp Glu Arg Leu Arg Arg Leu Val Gly Leu Trp Glu Pro

TCC AAA GCT CAA CAC TGC CAA CCC GGA GGT CAG GGA ATA CCT GAT AGG AGC

Ser Lys Ala Gln His Cys Gln Pro Gly Gly Gln Gly Ile Pro Asp Arg Ser

GGC CCT CCA CTG GAT AGA GTT CGG CTT TGA CGG CAT CAG GGT TGA TGT GCC

Gly Pro Pro Leu Asp Arg Val Arg Leu End Arg His Gln Gly End Cys Ala

GAA CGA AGT CCT CGA CCC GGG AAC GTT CTT CCC GGA GCT GAG AAA GGC AGT

Glu Arg Ser Pro Arg Pro Gly Asn Val Leu Pro Gly Ala Glu Lys Gly Ser

CAA GGA GAA AAA GCC GGA CGC ATA CCT CGT CGG TGA GAT ATG GAC GCT CTC

Gln Gly Glu Lys Ala Gly Arg Ile Pro Arg Arg End Asp Met Asp Ala Leu

CCC TGA GTG GGT GAA AGG AGA CCG CTT CGA CTC CCT CAT GAA CTA CGC CCT

Pro End Val Gly Glu Arg Arg Pro Leu Arg Leu Pro His Glu Leu Arg Pro

CGG GAG GGA CAT CCT CCT GAA CTA CGC GAA GGG CCT GCT CAG TGG AGA AAG

Arg Glu Gly His Pro Pro Glu Leu Arg Glu Gly Pro Ala Gln Trp Arg Lys

TGC AAT GAA AAT GAT GGG ACG TTA CTA TGC TTC CTA CGG CGA GAA CGT ATT

Cys Asn Glu Asn Asp Gly Thr Leu Leu Cys Phe Leu Arg Arg Glu Arg Ile

GCG ATG GGC TTC AAC CTC GTT GAT TCG CAC GAC ACT TCG AGG GTT CTC ACT

Ala Met Gly Phe Asn Leu Val Asp Ser His Asp Thr Ser Arg Val Leu Thr

GAT CTC GGT GGG GGG AGT CTC GGT GAC ACA CCG TCA AAC GAG TCA ATT CAG

Asp Leu Gly Gly Gly Ser Leu Gly Asp Thr Pro Ser Asn Glu Ser Ile Gln

AGA CTC AAG CTC CTC TCA ACG TCC TCT ATG CCC TGC CTG GAA CTC CGG TCA

Arg Leu Lys Leu Leu Ser Thr Ser Ser Met Pro Cys Leu Glu Leu Arg Ser

CCT TCC AGG GGA TGA GAG AGG ACT GCT CGG AGA CAA GGG GCA CTA CGA CGA

Pro Ser Arg Gly End Glu Arg Thr Ala Arg Arg Gln Gly Ala Leu Arg Arg

ACA GCG CTA CCC AAT ACA GTG GGA TAC TGT GAA CGA AGA CGT CCT GAA CCA

Thr Ala Leu Pro Asn Thr Val Gly Tyr Cys Glu Arg Arg Arg Pro Glu Pro

TTA CAG GGC ATT GGC GGA GCT CAG AAA AAG AGT TCC TGC ATT GAG GAG CAG

Leu Gln Gly Ile Gly Gly Ala Gln Lys Lys Ser Ser Cys Ile Glu Glu Gln

CGC AAT AAG GTT CTA CAC TGC CAA AGG CGG CGT TAT GGC CTT CTT CAG GGG

Arg Asn Lys Val Leu His Cys Gln Arg Arg Arg Tyr Gly Leu Leu Gln Gly

GCA TCA TGA CGA GGT TCT TGT CGT TGC CAA CAG CTG GAA GAA GCC AGC CCT

Ala Ser End Arg Gly Ser Cys Arg Cys Gln Gln Leu Glu Glu Ala Ser Pro

ACT AAA GCT TCC TGA GGG AGA GTG GAA AGT AAT CTG GCC TGA GAA TTT CAG

Thr Lys Ala Ser End Gly Arg Val Glu Ser Asn Leu Ala End Glu Phe Gln

CCC GGA ACT GCT TCG CGG CAA AGT TGA AGT GCC AGC CAT AGG GAT AAT CAT

Pro Gly Thr Ala Ser Arg Gln Ser End Ser Ala Ser His Arg Asp Asn His

CCT TGA GCG GAG TTG 1443

Pro End Ala Glu Leu

Bacillus thermoloeovorans (Clone # 68GC1) Glycosidase

1

ATG ACT GAA TTA TAT ATA AAA AAT CCC CTG ATC GAA CAG CGG GCA GAT CCC

Met Thr Glu Leu Tyr Ile Lys Asn Pro Leu Ile Glu Gln Arg Ala Asp Pro

TGG ATC TAT AAA CAT ACC GAT GGT TAT TAT TAC TTT ACC GGT TCC GTG CCG

Trp Ile Tyr Lys His Thr Asp Gly Tyr Tyr Tyr Phe Thr Gly Ser Val Pro

GAG TAC GAC CGA ATT GAG CTT AGA CGC TCG CAA ACG ATT CAA GGG CTT GCG

Glu Tyr Asp Arg Ile Glu Leu Arg Arg Ser Gln Thr Ile Gln Gly Leu Ala

GAT GCC GAA GGA ATT ACG ATC TGG CGC AAG CAT GAG TCA GGC CTG ATG AGT

Asp Ala Glu Gly Ile Thr Ile Trp Arg Lys His Glu Ser Gly Leu Met Ser

GCC AAC ATA TGG GCA CCC GAG ATT CAT TAT ATG GAT GGC AAA TGG TAT GTG

Ala Asn Ile Trp Ala Pro Glu Ile His Tyr Met Asp Gly Lys Trp Tyr Val

TAT TAC GCC GCT GCC CAT ACT TCA GAA ACG AGG GAC GGA TTG TTC GAT CAC

Tyr Tyr Ala Ala Ala His Thr Ser Glu Thr Arg Asp Gly Leu Phe Asp His

CGC ATG TTC GTA TTG GAG AAC GCT TCG GCG AAC CCG CTC GAA GGG GAA TGG

Arg Met Phe Val Leu Glu Asn Ala Ser Ala Asn Pro Leu Glu Gly Glu Trp

GTG GAG AAG GGG CAA GTG ATC ACG AAG TGG GAA TCT TTC GCC TTG GAC GCA
Val Glu Lys Gly Gln Val Ile Thr Lys Trp Glu Ser Phe Ala Leu Asp Ala

ACG ACG TTC GAG CAT AAA GGC AAA CGG TAC TAT GTA TGG GCT CAG AAA GAT
Thr Thr Phe Glu His Lys Gly Lys Arg Tyr Tyr Val Trp Ala Gln Lys Asp

CCG GGC ATT CCA GGC AAT TCC AAT CTG TAT ATC TCA TTG ATG GAA GAC CCG
Pro Gly Ile Pro Gly Asn Ser Asn Leu Tyr Ile Ser Leu Met Glu Asp Pro

TGG ACC CTG ACA GGG GAA CAG GTA TGC ATA TCG GTT CCC GAG TAC GAT TGG
Trp Thr Leu Thr Gly Glu Gln Val Cys Ile Ser Val Pro Glu Tyr Asp Trp

GAG AAG ATC GGG TAT CTT GTG AAT GAA GGG GCC GCC GTT CTT AAG CGA AAC
Glu Lys Ile Gly Tyr Leu Val Asn Glu Gly Ala Ala Val Leu Lys Arg Asn

GGG CGA ATA TTC ATG ACC TAT TCC GCG AGC GCC ACG GAC CAC AAC TAT GCG
Gly Arg Ile Phe Met Thr Tyr Ser Ala Ser Ala Thr Asp His Asn Tyr Ala

ATG GGG CTG CTG ACA GCC GAT GAA GAC AGT GAT TTG CTG AAT CCG AGC TCC
Met Gly Leu Leu Thr Ala Asp Glu Asp Ser Asp Leu Leu Asn Pro Ser Ser

TGG GTC AAG TCG CCT GTA CCT GTA TTT ACG ACA TCT GAA GCC AAT GGC CAA
Trp Val Lys Ser Pro Val Pro Val Phe Thr Thr Ser Glu Ala Asn Gly Gln

TAT GGT CCG GGG CAC AAC AGC TTC ACG ATT TCC GAG GAC GGC TTG CAG GAC

Tyr Gly Pro Gly His Asn Ser Phe Thr Ile Ser Glu Asp Gly Leu Gln Asp

ATT TTG GTA TAC CAT GCA AGA AGT TAC AAG GAG ATC GTC GGG ATC CAC TAT

Ile Leu Val Tyr His Ala Arg Ser Tyr Lys Glu Ile Val Gly Ile His Tyr

ATG ATC CGA ACC GTC ATA CGC GTG TAC AGG TCA TCC GAT GGA ACG AAG ACG

Met Ile Arg Thr Val Ile Arg Val Tyr Arg Ser Ser Asp Gly Thr Lys Thr

GAA CGC CGA ATT TCG GGG TGC CAA GAG CGG ATC ATG AAC CGG TCT CCA AGC

Glu Arg Arg Ile Ser Gly Cys Gln Glu Arg Ile Met Asn Arg Ser Pro Ser

CAT GAT GCC GAC TTT GTC ATT GGG GTT GTG ACC GGA AGG ATT AAC AAA CAT

His Asp Ala Asp Phe Val Ile Gly Val Val Thr Gly Arg Ile Asn Lys His

CAG ACC GAC TGA 1031

Gln Thr Asp END

Thermotoga maritima (Clone # 6GA2) Glycosidase

1

TTG AAT AAC ACC ATT CCA AGA TGG CGT GGT TTC AAC CTT CTG GAG GCC TTT

Leu Asn Asn Thr Ile Pro Arg Trp Arg Gly Phe Asn Leu Leu Glu Ala Phe

TCC ATT AAA AGT ACA GGA AAT TTT AAA GAG GAA GAT TTT TTG TGG ATG GCT

Ser Ile Lys Ser Thr Gly Asn Phe Lys Glu Glu Asp Phe Leu Trp Met Ala

CAG TGG GAC TTT AAT TTT GTT AGA ATC CCT ATG TGT CAT CTT CTC TGG TCA

Gln Trp Asp Phe Asn Phe Val Arg Ile Pro Met Cys His Leu Leu Trp Ser

GAC CGG GGC AAC CCA TTT ATT ATC AGA GAA GAT TTT TTT GAG AAA ATC GAT

Asp Arg Gly Asn Pro Phe Ile Ile Arg Glu Asp Phe Phe Glu Lys Ile Asp

CGT GTA ATT TTC TGG GGA GAG AAA TAT GGA ATA CAT ATA TGT ATT TCT CTT

Arg Val Ile Phe Trp Gly Glu Lys Tyr Gly Ile His Ile Cys Ile Ser Leu

CAC AGG GCA CCT GGC TAT TCT GTT AAC AAG GAA GTA GAA GAG AAA ACC AAT

His Arg Ala Pro Gly Tyr Ser Val Asn Lys Glu Val Glu Glu Lys Thr Asn

CTG TGG AAA GAT GAA ACA GCT CAA GAA GCG TTC ATT CAT CAC TGG TCT TTT

Leu Trp Lys Asp Glu Thr Ala Gln Glu Ala Phe Ile His His Trp Ser Phe

ATC GCA CGT CGT TAC AAA GGA ATT TCT TCC ACA CAC CTG AGT TTT AAC TTA
Ile Ala Arg Arg Tyr Lys Gly Ile Ser Ser Thr His Leu Ser Phe Asn Leu

ATA AAT GAG CCT CCA TTT CCT GAT CCA CAA ATC ATG AGT GTT GAA GAT CAC
Ile Asn Glu Pro Pro Phe Pro Asp Pro Gln Ile Met Ser Val Glu Asp His

AAC TCT CTT ATC AAG AGA ACT ATT ACA GAA ATT CGA AAA ATA GAT CCC GAA
Asn Ser Leu Ile Lys Arg Thr Ile Thr Glu Ile Arg Lys Ile Asp Pro Glu

AGA TTA ATT ATA ATA GAT GGA TTA GGC TAT GGG AAT ATT CCA GTG GAT GAT
Arg Leu Ile Ile Ile Asp Gly Leu Gly Tyr Gly Asn Ile Pro Val Asp Asp

TTA ACA ATT GAG AAT ACA GTG CAA TCA TGC AGA GGG TAC ATT CCC TTC AGT
Leu Thr Ile Glu Asn Thr Val Gln Ser Cys Arg Gly Tyr Ile Pro Phe Ser

GTT ACT CAT TAC AAA GCG GAA TGG GTG GAT AGT AAG GAC TTT CCT GTT CCT
Val Thr His Tyr Lys Ala Glu Trp Val Asp Ser Lys Asp Phe Pro Val Pro

GAG TGG CCA AAT GGA TGG CAT TTT GGG GAA TAC TGG AAC AGA GAA AAG TTA
Glu Trp Pro Asn Gly Trp His Phe Gly Glu Tyr Trp Asn Arg Glu Lys Leu

TTG GAA CAT TAT TTA ACG TGG ATA AAA CTC AGA CAA AAA GGA ATA GAA GTA
Leu Glu His Tyr Leu Thr Trp Ile Lys Leu Arg Gln Lys Gly Ile Glu Val

TTC TGT GGA GAA ATG GGA GCT TAC AAC AAA ACA CCT CAC GAT GTG GTT TTA

Phe Cys Gly Glu Met Gly Ala Tyr Asn Lys Thr Pro His Asp Val Val Leu

AAA TGG CTT GAA GAT CTT TTA GAA ATT TTT AAA ACT TTG AAC ATA GGG TTT

Lys Trp Leu Glu Asp Leu Leu Glu Ile Phe Lys Thr Leu Asn Ile Gly Phe

GCC TTA TGG AAT TTT AGA GGT CCT TTT GGT ATT TTA GAT TCG GAA AGG AAA

Ala Leu Trp Asn Phe Arg Gly Pro Phe Gly Ile Leu Asp Ser Glu Arg Lys

GAC GTT GAA TAC GAA GAA TGG TAT GGA CAT AAA CTG GAT AGG AAA ATG TTG

Asp Val Glu Tyr Glu Glu Trp Tyr Gly His Lys Leu Asp Arg Lys Met Leu

GAA CTA TTG AGA AAA TAT TAG 990

Glu Leu Leu Arg Lys Tyr End

Thermotoga maritima MSB8 (Clone # 6GC17) Glycosidase

1

ATG CTC TCA GAG ATT GTT CCG TAT ACT GTT CTG AGA AGA GAA AGA ATA GAA

Met Leu Ser Glu Ile Val Pro Tyr Thr Val Leu Arg Arg Glu Arg Ile Glu

AGC TGG ATT TTC TCC GAT GAT GCT GTT GAG AGA ATC GTG GAT CCT TCC TTC

Ser Trp Ile Phe Ser Asp Asp Ala Val Glu Arg Ile Val Asp Pro Ser Phe

GAA TGG GAC TTC AGC TCC GCT CCC GTC CGG TTC AGG AAA GAG CTA GAG CCT

Glu Trp Asp Phe Ser Ser Ala Pro Val Arg Phe Arg Lys Glu Leu Glu Pro

TTC TCC GTC GCT GGA GAG CAG AGG GCC TAC CTG AAA CTC TGG TTC GGT GGT

Phe Ser Val Ala Gly Glu Gln Arg Ala Tyr Leu Lys Leu Trp Phe Gly Gly

GAA ACA CTC GTT CTG ATA GAT GGG AAG CCT TAC GGT GAG ATC AAC GAG TAT

Glu Thr Leu Val Leu Ile Asp Gly Lys Pro Tyr Gly Glu Ile Asn Glu Tyr

CAT AGG ATG TTG AAC ATC ACC CCC CTT GCT GAT GGA AAA CCA CAC ACG ATA

His Arg Met Leu Asn Ile Thr Pro Leu Ala Asp Gly Lys Pro His Thr Ile

GAA GCT CAG GTG ATG CCA AGG GGT CTC TTT GGA AAA CCA GAA AAG CCG GTG

Glu Ala Gln Val Met Pro Arg Gly Leu Phe Gly Lys Pro Glu Lys Pro Val

TTC ACG GAA GCT TTC TTC ATC GTC GTT GAT GAA GCA CTG ATG AAG GTG GTG

Phe Thr Glu Ala Phe Phe Ile Val Val Asp Glu Ala Leu Met Lys Val Val

AAA ACT CTC GAA CTC ACT ATA AAA ACG GCA GAA GTG ATA GAA GAC GAG TCG

Lys Thr Leu Glu Leu Thr Ile Lys Thr Ala Glu Val Ile Glu Asp Glu Ser

CTT TCT AAG AAA CTT CTG GAC ATC TCC GAG GAG TTT CTC TCG AAA GTA TGG

Leu Ser Lys Lys Leu Leu Asp Ile Ser Glu Glu Phe Leu Ser Lys Val Trp

ATC CCA AGA GAC ACA GGT ACC TAT CTG ATG ACA GCA CTG GAG GAT CCG GGA

Ile Pro Arg Asp Thr Gly Thr Tyr Leu Met Thr Ala Leu Glu Asp Pro Gly

ATA AAA GAT GAG ATC AAA AAC ACC TGG AAC ACA CCG GAG TTC AAA GAG TTC

Ile Lys Asp Glu Ile Lys Asn Thr Trp Asn Thr Pro Glu Phe Lys Glu Phe

ACA GGT GTG AAG CTT CCT GAA GAG TTG AGA AAT CAG ATT CTG GAA GAG TTC

Thr Gly Val Lys Leu Pro Glu Glu Leu Arg Asn Gln Ile Leu Glu Glu Phe

GAA AAA TTC AAA GAA AAG CTG GAT AGA ATA AGA AAA AAC CAT CCG GGT TTT

Glu Lys Phe Lys Glu Lys Leu Asp Arg Ile Arg Lys Asn His Pro Gly Phe

GGA ACG ATT CAC CTT GTG GGG CAC GCG CAC ATA GAC TAC GCC TGG CTC TGG

Gly Thr Ile His Leu Val Gly His Ala His Ile Asp Tyr Ala Trp Leu Trp

CCA GTT GAG GAG ACG AAG AGA AAG ATC CTA CGC ACT TTC GCA AAC TCT GTG

Pro Val Glu Glu Thr Lys Arg Lys Ile Leu Arg Thr Phe Ala Asn Ser Val

TTG CTC TCT AAG CTT TAT CCG GAG TTC GTT TAC ACT CAG TCT TCT GCT CAG

Leu Leu Ser Lys Leu Tyr Pro Glu Phe Val Tyr Thr Gln Ser Ser Ala Gln

ATG TAC GAG GAT CTC AAG CAA AAT TCA CCA GAG CTT TTC GAG GAA GTG AGA

Met Tyr Glu Asp Leu Lys Gln Asn Ser Pro Glu Leu Phe Glu Glu Val Arg

AAG CTC GTA GAA GAG GGG AGA TGG GAG CCA GTC GGT GGC ATG TGG GTG GAG

Lys Leu Val Glu Glu Gly Arg Trp Glu Pro Val Gly Gly Met Trp Val Glu

TCG GAC TGC AAC GTT CCA TCG ATA GAG TCG CTT GTG AGA CAG TTC TAC TAT

Ser Asp Cys Asn Val Pro Ser Ile Glu Ser Leu Val Arg Gln Phe Tyr Tyr

GGG CAA AAA TTC TTC GAA AGA GAA TTC GGG AAA AAG AGC AAG GTG TGC TGG

Gly Gln Lys Phe Phe Glu Arg Glu Phe Gly Lys Lys Ser Lys Val Cys Trp

CTT CCG GAT GTG TTT GGG TTT TCC TGG GTG CTT CCC CAA ATT CTG AAA GAA

Leu Pro Asp Val Phe Gly Phe Ser Trp Val Leu Pro Gln Ile Leu Lys Glu

GCC GGG ATA AAA TAC TTC GTC ACC ACG AAA CTC AAC TGG AAC GAC ACG AAC

Ala Gly Ile Lys Tyr Phe Val Thr Thr Lys Leu Asn Trp Asn Asp Thr Asn

GAG TTT CCG TAC GAT CTG TGC CGC TGG AGG GGA ATA GAT GGA TCC GAA GTG
Glu Phe Pro Tyr Asp Leu Cys Arg Trp Arg Gly Ile Asp Gly Ser Glu Val

ATC TAT TTC AGT TTC AAA AAT CCC AAC GAG GGG TAC AAC GGA AAG ATA GAT
Ile Tyr Phe Ser Phe Lys Asn Pro Asn Glu Gly Tyr Asn Gly Lys Ile Asp

CCC GAT ACG GTC TAC AAA ACC TGG AAG AAC TTC AGG CAG AAA GAT CTC ACA
Pro Asp Thr Val Tyr Lys Thr Trp Lys Asn Phe Arg Gln Lys Asp Leu Thr

AAC AGA GTT CTT CTT TCG TTC GGA CAC GGT GAT GGT GGT GGC GGT CCA ACC
Asn Arg Val Leu Leu Ser Phe Gly His Gly Asp Gly Gly Gly Gly Pro Thr

GAA GAG ATG CTG GAA AAT TAC GAG GTT CTG AAG GAT TTC CCT GGA CTA CCG
Glu Glu Met Leu Glu Asn Tyr Glu Val Leu Lys Asp Phe Pro Gly Leu Pro

CAC CTT GAA ATG GGA ACT GTG GAA GAA TTT TTC AAG AAG GTG GAG ATC GAC
His Leu Glu Met Gly Thr Val Glu Glu Phe Phe Lys Lys Val Glu Ile Asp

GAA GAA CTC CCT GTG TGG GAC GGA GAG CTT TAC CTT GAA CTT CAC AGG GGA
Glu Glu Leu Pro Val Trp Asp Gly Glu Leu Tyr Leu Glu Leu His Arg Gly

ACC TAC ACT TCT CAG TTC AGG ACA AAG AAA CTT CAC AAA GAA GCG GAA GAC

Thr Tyr Thr Ser Gln Phe Arg Thr Lys Lys Leu His Lys Glu Ala Glu Asp

AGT CTT TAT CTT GCA GAG TTG ATC TCG GCT TTC ACG GAT AAA GAT TTT TCG

Ser Leu Tyr Leu Ala Glu Leu Ile Ser Ala Phe Thr Asp Lys Asp Phe Ser

GAC GAA ATA GAC GAA CTC TGG AAG ATT CTG TTG AGA AAC GAA TTT CAC GAT

Asp Glu Ile Asp Glu Leu Trp Lys Ile Leu Leu Arg Asn Glu Phe His Asp

ATT CTA CCT GGA TCT TCT ATA AAG GAA GTC TAT GAA GAT ACA GAA AAA GAG

Ile Leu Pro Gly Ser Ser Ile Lys Glu Val Tyr Glu Asp Thr Glu Lys Glu

CTC AGA CAT GTG ATA GAA AAA TCA AAA GAC ATC GTT ATC GAA TCT CTC AAA

Leu Arg His Val Ile Glu Lys Ser Lys Asp Ile Val Ile Glu Ser Leu Lys

GTT CTT TCC TCT GAG AAC AAA GAT GTT CTA ACC ATT TTG AAC GCT TCA TCG

Val Leu Ser Ser Glu Asn Lys Asp Val Leu Thr Ile Leu Asn Ala Ser Ser

TTT CCA AAG AAG TGT CTT TTC TTC CTC AAC GAA GAT CTC GCG ATT TCC TTT

Phe Pro Lys Lys Cys Leu Phe Phe Leu Asn Glu Asp Leu Ala Ile Ser Phe

GAA GGA GAA GCA CTC TTG AAA CAG AAA ACT CAC GAT GGA AGG TAT GTG TAC

Glu Gly Glu Ala Leu Leu Lys Gln Lys Thr His Asp Gly Arg Tyr Val Tyr

TTC ATA GAC AGG GAG ATT CCT CCG TTC ACG AAA GTA GAA CTG AAA GTT CGC

Phe Ile Asp Arg Glu Ile Pro Pro Phe Thr Lys Val Glu Leu Lys Val Arg

AAA GCC ACG TCT GAG GAA ACT CCA AGT GAG TTG AGA GAA ACA AAC ATC ATG

Lys Ala Thr Ser Glu Glu Thr Pro Ser Glu Leu Arg Glu Thr Asn Ile Met

GAG AAC GAA TTT CTC AGG GTG CAC GTC AAC GAT GAC GGA ACA ATT CAA ATC

Glu Asn Glu Phe Leu Arg Val His Val Asn Asp Asp Gly Thr Ile Gln Ile

TAC GAC AAA GAA CTG GAC AGG TAC GTT TTC GAA GAG AAG GGA AAC ATC TTG

Tyr Asp Lys Glu Leu Asp Arg Tyr Val Phe Glu Glu Lys Gly Asn Ile Leu

AAA CTT CAT AAA AAC ATC CCT GCT TAC TGG GAC AAC TGG GAT ATC GCA GAA

Lys Leu His Lys Asn Ile Pro Ala Tyr Trp Asp Asn Trp Asp Ile Ala Glu

AAC GTG GAA AAG ACA GGA TAT ACC CTG AGG GCG AAA AAC ATA GAA AAA ATA

Asn Val Glu Lys Thr Gly Tyr Thr Leu Arg Ala Lys Asn Ile Glu Lys Ile

GAG TCT GGC CCT GTT CGA GAA GTG ATC CGT GTT GAA CAT GAA TCA GAA GGA

Glu Ser Gly Pro Val Arg Glu Val Ile Arg Val Glu His Glu Ser Glu Gly

AGC AGG ATC ACG CAG CAT TAC ATC CTT TAC AGA AAG AGT AGA AGG CTC GAT

Ser Arg Ile Thr Gln His Tyr Ile Leu Tyr Arg Lys Ser Arg Arg Leu Asp

ATA GAA ACG AAG GTA GAC TGG CAC ACA AGG CGT GCG CTT CTC AGA GCC TAC

Ile Glu Thr Lys Val Asp Trp His Thr Arg Arg Ala Leu Leu Arg Ala Tyr

TTC CCA ACA ACT GTT CTG TCG AGA AAG GCT AGG TTC GAT ATC TCC GGT GGT

Phe Pro Thr Thr Val Leu Ser Arg Lys Ala Arg Phe Asp Ile Ser Gly Gly

TTC ATC GAA AGG CCC ACA CAC AGA AAC ACC AGT TTC GAA CAG GCG CGT TTC

Phe Ile Glu Arg Pro Thr His Arg Asn Thr Ser Phe Glu Gln Ala Arg Phe

GAG GTG CCG TTT CAC AGG TGG ATG GAT CTT TCC CAG ACA GAC TTC GGC GTG

Glu Val Pro Phe His Arg Trp Met Asp Leu Ser Gln Thr Asp Phe Gly Val

TCC ATT CTG AAC GAC GGA AAA TAC GGT GGC AGT GTT CAT CAG GGT ATC ATG

Ser Ile Leu Asn Asp Gly Lys Tyr Gly Gly Ser Val His Gln Gly Ile Met

GCG CTT TCA CTG ATA AAA GCG GGT ATT TTC CCC GAT TTT CTC TGT GAC GAA

Ala Leu Ser Leu Ile Lys Ala Gly Ile Phe Pro Asp Phe Leu Cys Asp Glu

GGC GAA CAC ACT TTC ACC TAT TCT GTC TAC GTA CAC CCT GGA GAC AGC TTG

Gly Glu His Thr Phe Thr Tyr Ser Val Tyr Val His Pro Gly Asp Ser Leu

AGA GAT GTT GTA AAA GGA TCA GAA GAT CTC AAC AGA TCT TTC ATC GTT CAT

Arg Asp Val Val Lys Gly Ser Glu Asp Leu Asn Arg Ser Phe Ile Val His

CGC GGG GTG TTG AAC CTC CCC TCT CCT TTA CTG GAG ATC TCT CCT CAA AAC

Arg Gly Val Leu Asn Leu Pro Ser Pro Leu Leu Glu Ile Ser Pro Gln Asn

TTC CGT CTC ACC TCA CTG AGA AGG GTG AAG GAC AAA ATT GTT TTG AGG CTT

Phe Arg Leu Thr Ser Leu Arg Arg Val Lys Asp Lys Ile Val Leu Arg Leu

GTT GAG ATT TTC GGA ACA TCA GGG AAA CTT TCC ATT AAA CTC CCA TGG CAT

Val Glu Ile Phe Gly Thr Ser Gly Lys Leu Ser Ile Lys Leu Pro Trp His

GGT GAA ATC TAT CAG ACG AAC GTT CTG GAA GAG AAA AAA CAG AAA GTC ACC

Gly Glu Ile Tyr Gln Thr Asn Val Leu Glu Glu Lys Lys Gln Lys Val Thr

TTC CCA GTG GTT TAC CAT CCG TTC AAG ATC TAC ACT TTT GTT GTA GAA GGT

Phe Pro Val Val Tyr His Pro Phe Lys Ile Tyr Thr Phe Val Val Glu Gly

TGA 3011

END

Thermotoga maritima MSB8 (Clone # 6GC18) Glycosidase

1
ATG GAA CTG TAC AGG GAT CCT TCG CAA CCC ATC GAA GTG AGA GTG AGA GAT
Met Glu Leu Tyr Arg Asp Pro Ser Gln Pro Ile Glu Val Arg Val Arg Asp

CTT CTT TCC AGA ATG ACG CTG GAA GAG AAA GTG GCC CAG CTT GGG TCT GTC
Leu Leu Ser Arg Met Thr Leu Glu Glu Lys Val Ala Gln Leu Gly Ser Val

TGG GGT TAC GAA CTG ATA GAC GAG AGG GGA AAG TTC AGT AGA GAA AAA GCA
Trp Gly Tyr Glu Leu Ile Asp Glu Arg Gly Lys Phe Ser Arg Glu Lys Ala

AAA GAA CTC CTC AAA AAT GGT ATA GGC CAG ATC ACA AGG CCT GGT GGA TCA
Lys Glu Leu Leu Lys Asn Gly Ile Gly Gln Ile Thr Arg Pro Gly Gly Ser

ACG AAC CTT GAA CCT CAA GAA GCC GCG GAA CTT GTG AAC GAA ATA CAG AGA
Thr Asn Leu Glu Pro Gln Glu Ala Ala Glu Leu Val Asn Glu Ile Gln Arg

TTT CTT GTG GAA GAA ACA CGC CTT GGA ATT CCT GCG ATG ATA CAC GAA GAA
Phe Leu Val Glu Glu Thr Arg Leu Gly Ile Pro Ala Met Ile His Glu Glu

TGT CTC ACC GGT TAC ATG GGA CTT GGA GGA ACC AAC TTC CCT CAG GCG ATA
Cys Leu Thr Gly Tyr Met Gly Leu Gly Gly Thr Asn Phe Pro Gln Ala Ile

GCA ATG GCG AGT ACA TGG GAT CCA GAT CTC ATA GAA AAA ATG ACC ACC GCC

Ala Met Ala Ser Thr Trp Asp Pro Asp Leu Ile Glu Lys Met Thr Thr Ala

GTC AGA GAG GAT ATG AGA AAG ATA GGG GCA CAT CAG GGT CTC GCA CCT GTT

Val Arg Glu Asp Met Arg Lys Ile Gly Ala His Gln Gly Leu Ala Pro Val

CTG GAT GTC GCA AGA GAT CCA AGG TGG GGG AGA ACA GAA GAG ACG TTC GGA

Leu Asp Val Ala Arg Asp Pro Arg Trp Gly Arg Thr Glu Glu Thr Phe Gly

GAA TCT CCC TAT CTG GTG GCG AGG ATG GGA GTC TCT TAC GTG AAA GGC CTC

Glu Ser Pro Tyr Leu Val Ala Arg Met Gly Val Ser Tyr Val Lys Gly Leu

CAG GGG GAA GAT ATC AAA AAA GGT GTC GTT GCC ACA GTG AAA CAC TTC GCC

Gln Gly Glu Asp Ile Lys Lys Gly Val Val Ala Thr Val Lys His Phe Ala

GGA TAC AGC GCT TCT GAA GGT GGA AAG AAC TGG GCA CCA ACG AAC ATT CCG

Gly Tyr Ser Ala Ser Glu Gly Gly Lys Asn Trp Ala Pro Thr Asn Ile Pro

GAG AGG GAA TTC AAA GAG GTC TTT CTC TTT CCG TTC GAA GCG GCC GTT AAA

Glu Arg Glu Phe Lys Glu Val Phe Leu Phe Pro Phe Glu Ala Ala Val Lys

GAA GCG AAT GTG CTT TCT GTG ATG AAC TCC TAC AGC GAA ATA GAC GGT GTC

Glu Ala Asn Val Leu Ser Val Met Asn Ser Tyr Ser Glu Ile Asp Gly Val

CCA TGT GCA GCG AAC AGG AAA CTC CTC ACA GAC ATT CTC AGA AAA GAC TGG

Pro Cys Ala Ala Asn Arg Lys Leu Leu Thr Asp Ile Leu Arg Lys Asp Trp

GGA TTC GAA GGA ATC GTC GTT TCT GAC TAT TTT GCT GTG AAA GTT CTG GAA

Gly Phe Glu Gly Ile Val Val Ser Asp Tyr Phe Ala Val Lys Val Leu Glu

GAT TAT CAC AGA ATA GCA AGG GAT AAG TCA GAA GCC GCA AGA CTC GCA CTT

Asp Tyr His Arg Ile Ala Arg Asp Lys Ser Glu Ala Ala Arg Leu Ala Leu

GAA GCG GGG ATA GAT GTT GAA CTT CCG AAG ACA GAA TGT TAT CAA TAT TTG

Glu Ala Gly Ile Asp Val Glu Leu Pro Lys Thr Glu Cys Tyr Gln Tyr Leu

AAA GAC CTT GTT GAA AAA GGC ATC ATC TCC GAA GCT TTG ATC GAC GAG GCA

Lys Asp Leu Val Glu Lys Gly Ile Ile Ser Glu Ala Leu Ile Asp Glu Ala

GTC ACC AGG GTG CTG AGG CTG AAG TTC ATG CTC GGG CTC TTC GAA AAT CCC

Val Thr Arg Val Leu Arg Leu Lys Phe Met Leu Gly Leu Phe Glu Asn Pro

TAC GTT GAG GTG GAA AAA GCA AAG ATA GAA AGT CAC AGA GAC ATC GCA CTC

Tyr Val Glu Val Glu Lys Ala Lys Ile Glu Ser His Arg Asp Ile Ala Leu

GAG ATA GCA AGG AAA TCC ATT ATC CTT CTC AAG AAT GAT GGA ATT CTG CCT

Glu Ile Ala Arg Lys Ser Ile Ile Leu Leu Lys Asn Asp Gly Ile Leu Pro

CTT CAG AAA AAC AAA AAA GTT GCC CTG ATC GGA CCG AAC GCG GGT GAG GTG

Leu Gln Lys Asn Lys Lys Val Ala Leu Ile Gly Pro Asn Ala Gly Glu Val

AGA AAT CTC CTC GGA GAT TAC ATG TAC CTT GCA CAC ATA AGG GCT CTC CTC

Arg Asn Leu Leu Gly Asp Tyr Met Tyr Leu Ala His Ile Arg Ala Leu Leu

GAC AAC ATA GAC GAC GTC TTT GGA AAT CCT CAG ATC CCG AGA GAA AAC TAC

Asp Asn Ile Asp Asp Val Phe Gly Asn Pro Gln Ile Pro Arg Glu Asn Tyr

GAA AGA CTG AAG AAG AGC ATA GAA GAA CAT ATG AAG AGC ATT CCG AGT GTT

Glu Arg Leu Lys Lys Ser Ile Glu Glu His Met Lys Ser Ile Pro Ser Val

CTC GAT GCC TTC AAA GAA GAA GGG ATC GAA TTC GAA TAT GCA AAA GGC TGT

Leu Asp Ala Phe Lys Glu Glu Gly Ile Glu Phe Glu Tyr Ala Lys Gly Cys

GAA GTG ACA GGG GAA GAC AGA AGC GGT TTC GAA GAG GCG ATA GAA ATT GCA

Glu Val Thr Gly Glu Asp Arg Ser Gly Phe Glu Glu Ala Ile Glu Ile Ala

AAG AAA TCC GAC GTT GCC ATC GTT GTC GTA GGG GAC AAA TCT GGA CTC ACC

Lys Lys Ser Asp Val Ala Ile Val Val Val Gly Asp Lys Ser Gly Leu Thr

CTT GAC TGC ACA ACC GGT GAG TCC AGA GAC ATG GCA AAC CTC AAG CTT CCA

Leu Asp Cys Thr Thr Gly Glu Ser Arg Asp Met Ala Asn Leu Lys Leu Pro

GGA GTC CAG GAA GAA CTC GTC CTC GAA GTT GCA AAG ACA GGA AAA CCC GTC

Gly Val Gln Glu Glu Leu Val Leu Glu Val Ala Lys Thr Gly Lys Pro Val

GTT CTT GTC CTC ATC ACG GGA AGA CCC TAT TCA CTC AAA AAC GTC GTC GAC

Val Leu Val Leu Ile Thr Gly Arg Pro Tyr Ser Leu Lys Asn Val Val Asp

AAG GTG AAC GCG ATC CTT CAG GTG TGG CTT CCT GGA GAA GCG GGA GGA AGA

Lys Val Asn Ala Ile Leu Gln Val Trp Leu Pro Gly Glu Ala Gly Gly Arg

GCG ATC GTT GAC ATC ATC TAT GGA AAG GTG AAT CCC TCT GGA AAA CTC CCG

Ala Ile Val Asp Ile Ile Tyr Gly Lys Val Asn Pro Ser Gly Lys Leu Pro

ATC AGC TTT CCA AGA AGC GCT GGT CAG ATT CCT GTC TTC CAC TAC GTC AAA

Ile Ser Phe Pro Arg Ser Ala Gly Gln Ile Pro Val Phe His Tyr Val Lys

CCA TCC GGG GGA AGG TCT CAC TGG CAC GGA GAC TAC GTG GAT GAG AGC ACA

Pro Ser Gly Gly Arg Ser His Trp His Gly Asp Tyr Val Asp Glu Ser Thr

AAG CCT CTC TTC CCG TTT GGG CAC GGT TTG TCT TAC ACG AAG TTC GAG TAC

Lys Pro Leu Phe Pro Phe Gly His Gly Leu Ser Tyr Thr Lys Phe Glu Tyr

AGC AAC CTC AGA ATC GAG CCG AAG GAA GTG CCA CCG GCC GGC GAA GTG GTG
Ser Asn Leu Arg Ile Glu Pro Lys Glu Val Pro Pro Ala Gly Glu Val Val

ATA AAG GTG GAC GTG GAA AAC ATC GGA GAC AGA GAC GGA GAC GAG GTG GTT
Ile Lys Val Asp Val Glu Asn Ile Gly Asp Arg Asp Gly Asp Glu Val Val

CAA CTT TAC ATC GGT CGT GAG TTT GCA AGC GTC ACA AGG CCT GTG AAA GAG
Gln Leu Tyr Ile Gly Arg Glu Phe Ala Ser Val Thr Arg Pro Val Lys Glu

CTG AAG GGC TTC AAG AGG GTT TCT TTG AAG GCG AAA GAG AAG AAG ACT GTT
Leu Lys Gly Phe Lys Arg Val Ser Leu Lys Ala Lys Glu Lys Lys Thr Val

GTG TTC AGG CTT CAC ATG GAC GTG CTC GCC TAC TAC AAC AGA GAC ATG AAA
Val Phe Arg Leu His Met Asp Val Leu Ala Tyr Tyr Asn Arg Asp Met Lys

CTC GTG GTT GAA CCC GGT GAG TTC AAA GTG ATG GTG GGA AGC TCT TCT GAA
Leu Val Val Glu Pro Gly Glu Phe Lys Val Met Val Gly Ser Ser Ser Glu

GAC ATC AGA CTC ACA GGT TCT TTC TCC GTC GTC GGT GAA AAA AGA GAA GTG
Asp Ile Arg Leu Thr Gly Ser Phe Ser Val Val Gly Glu Lys Arg Glu Val

GTG GGA ATG AGG AAA TTC TTC ACG GAA GCC TGC GAG GAG TGA 2336
Val Gly Met Arg Lys Phe Phe Thr Glu Ala Cys Glu Glu END

Thermotoga maritima MSB8 (Clone # 6GP2) Glycosidase

1

ATG GGG ATT GGT GGC GAC GAC TCC TGG AGC CCG TCA GTA TCG GCG GAA TTC

Met Gly Ile Gly Gly Asp Asp Ser Trp Ser Pro Ser Val Ser Ala Glu Phe

CTT TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC

Leu Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe

GTG AAA GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC

Val Lys Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe

ATT GGA AGC AAC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC

Ile Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp

AGT GTT CTG GAG AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG

Ser Val Leu Glu Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp

GGT TTC CTC GAC GGG GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT

Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His

CCT GAG CCC GGT GTT TTC GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC

Pro Glu Pro Gly Val Phe Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser

GGT TTC GAA AGA CTC GAC TAC ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA

Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala Lys Glu Leu Gly Ile

AAA CTT GTC ATT GTT CTT GTG AAC AAC TGG GAC GAC TTC GGT GGA ATG AAC

Lys Leu Val Ile Val Leu Val Asn Asn Trp Asp Asp Phe Gly Gly Met Asn

CAG TAC GTG AGG TGG TTT GGA GGA ACC CAT CAC GAC GAT TTC TAC AGA GAT

Gln Tyr Val Arg Trp Phe Gly Gly Thr His His Asp Asp Phe Tyr Arg Asp

GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC GTC TCC TTT CTC GTA AAC CAT

Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser Phe Leu Val Asn His

GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA GAG CCC ACC ATC ATG GCC

Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu Pro Thr Ile Met Ala

TGG GAG CTT GCA AAC GAA CCG CGC TGT GAG ACG GAC AAA TCG GGG AAC ACG

Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp Lys Ser Gly Asn Thr

CTC GTT GAG TGG GTG AAG GAG ATG AGC TCC TAC ATA AAG AGT CTG GAT CCC

Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys Ser Leu Asp Pro

AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC TAC GAA GGA

Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn Tyr Glu Gly

TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG TCC GGT

Phe Lys Pro Tyr Gly Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp Ser Gly

GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG TTC

Val Asp Trp Lys Lys Leu Leu Ser Ile Glu Thr Val Asp Phe Gly Thr Phe

CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG

His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp

GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA

Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys

CCC GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA

Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg

ACG GCC ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT

Thr Ala Ile Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp

GGA GCG ATG TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC

Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp

GAG AGA GGG TAC TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC

Glu Arg Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp

AGT CCA GAA GCG GAA CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT

Ser Pro Glu Ala Glu Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly

GAA GAC ATA AGA GAA GAC ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG

Glu Asp Ile Arg Glu Asp Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met

GAG ATC AAA AAG ACC GTG GAA GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC

Glu Ile Lys Lys Thr Val Glu Val Arg Ala Gly Val Phe Asp Tyr Ser Asn

ACG TTT GAA AAG TTG TCT GTC AAA GTC GAA GAT CTG GTT TTT GAA AAT GAG

Thr Phe Glu Lys Leu Ser Val Lys Val Glu Asp Leu Val Phe Glu Asn Glu

ATA GAG CAT CTC GGA TAC GGA ATT TAC GGC TTT GAT CTC GAC ACA ACC CGG

Ile Glu His Leu Gly Tyr Gly Ile Tyr Gly Phe Asp Leu Asp Thr Thr Arg

ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT GAA GGC CAC TTT CAG GGA AAA

Ile Pro Asp Gly Glu His Glu Met Phe Leu Glu Gly His Phe Gln Gly Lys

ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG AAC GAA GCA CGG TAC GTG

Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val Asn Glu Ala Arg Tyr Val

CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG GTG AAA AAC TGG TGG

Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu Val Lys Asn Trp Trp

AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC ATT GAA TGG AAC

Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp Ile Glu Trp Asn

GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTG CCC GGA AAG

Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu Pro Gly Lys

AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC TCA GAA

Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu Ser Glu

TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC AAG

Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu Lys

GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC

Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly

CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC

Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly

GGA AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG

Gly Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala

GGG GTG AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT

Gly Val Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp

GGA CCG ATT TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG

Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met

TGA 2042

END

Polyangium brachysporum (Clone # 78GA1) Glycosidase

1

ATG TTC CTG CAT CCG AGG GGT CGC ATG ACC CGC CTA GCG CTC GGC TGT GCC

Met Phe Leu His Pro Arg Gly Arg Met Thr Arg Leu Ala Leu Gly Cys Ala

GTG CTG TGT CTG GCC GTC GCA GGC TGC GGT GGT GGT GAT GAC GAC GGC GAC

Val Leu Cys Leu Ala Val Ala Gly Cys Gly Gly Gly Asp Asp Asp Gly Asp

GAC AAC GGC ACC GCC CCC CAG CCC GCA CCT GGT CAA CCC GAG CCC CCG ACT

Asp Asn Gly Thr Ala Pro Gln Pro Ala Pro Gly Gln Pro Glu Pro Pro Thr

GAC ACC GTG CTG AAA GAC TGG CCT CGC ATC AAC AGC AGC ATC ACC GCC GAC

Asp Thr Val Leu Lys Asp Trp Pro Arg Ile Asn Ser Ser Ile Thr Ala Asp

GCA GCG ATC GAA AGC CGC GTC AAC TCA CTC GTC GCG GCG ATG ACG CTG GAA

Ala Ala Ile Glu Ser Arg Val Asn Ser Leu Val Ala Ala Met Thr Leu Glu

GAA AAA GTC GGC CAG ATG ACG CAG GTC GAA ATC CAG GAG GTG ACG CCG GAG

Glu Lys Val Gly Gln Met Thr Gln Val Glu Ile Gln Glu Val Thr Pro Glu

GAG ATC CGG CAG TAC CAC ATC GGC TCC GTG CTC AAC GGC GGT GGT TCG TTC

Glu Ile Arg Gln Tyr His Ile Gly Ser Val Leu Asn Gly Gly Gly Ser Phe

CCG AAG CAG GAC AAG GGC GCG GCG GTG ACC GAC TGG CTG GCG GTG GCC GAC
Pro Lys Gln Asp Lys Gly Ala Ala Val Thr Asp Trp Leu Ala Val Ala Asp

GCC TTG TGG GCC GCG TCG ATG GAT CCC GCC AAG CCG CGG CGC ATC CCG CTC
Ala Leu Trp Ala Ala Ser Met Asp Pro Ala Lys Pro Arg Arg Ile Pro Leu

ATC TGG GGC ACC GAC GCC GTC CAC GGC CAC AAC AAC GTC AAG GGC GCG ACC
Ile Trp Gly Thr Asp Ala Val His Gly His Asn Asn Val Lys Gly Ala Thr

ATC TTC CCG CAC AAC ATC GGC CTG GGC GCC GCG CGC GAC CCC GAC TTG GTC
Ile Phe Pro His Asn Ile Gly Leu Gly Ala Ala Arg Asp Pro Asp Leu Val

GCC CGC ATC GGC GCC GCC ACG GCG CTG GAA GTG GCA CGC ACC GGC ATC GAC
Ala Arg Ile Gly Ala Ala Thr Ala Leu Glu Val Ala Arg Thr Gly Ile Asp

TGG GTG TTC GCG CCA ACG CTG GCG GTC GTG CGC GAC GAC CGC TGG GGC CGC
Trp Val Phe Ala Pro Thr Leu Ala Val Val Arg Asp Asp Arg Trp Gly Arg

AGC TAC GAA GGC TAT TCG GAA GAC CCC GAA ATC GTC GTC TCC TAT GCC GGC
Ser Tyr Glu Gly Tyr Ser Glu Asp Pro Glu Ile Val Val Ser Tyr Ala Gly

AAG ATG GTC GAA GGC CTG CAG GGC CGA TTG GCG CAG GAC GCG AAG GCC AAC
Lys Met Val Glu Gly Leu Gln Gly Arg Leu Ala Gln Asp Ala Lys Ala Asn

GAG AAG GTG GTG GCC ACC GCC AAG CAT TTC GTC GGC GAC GGC GGC ACC GAC

Glu Lys Val Val Ala Thr Ala Lys His Phe Val Gly Asp Gly Gly Thr Asp

CAG GGC AAG GAC CAG GGG GTC ACC GGG GTC ACC GAG CGC GAC CTG TTG AAC

Gln Gly Lys Asp Gln Gly Val Thr Arg Val Thr Glu Arg Asp Leu Leu Asn

GTC CAT GCG CGC GGC TAC ATC CCC GCG CTC GAG GCG GGC GCG CAA ACC GTG

Val His Ala Arg Gly Tyr Ile Pro Ala Leu Glu Ala Gly Ala Gln Thr Val

ATG GCC TCC TTC AAC AGC TGG CAG GAC CCG TCG CAG GGC GAG GGC GCC AAG

Met Ala Ser Phe Asn Ser Trp Gln Asp Pro Ser Gln Gly Glu Gly Ala Lys

GCC TTC AAG ATG CAT GGC AGC CGC TAC CTG CTC ACC GAG GCC CTC AAG CAG

Ala Phe Lys Met His Gly Ser Arg Tyr Leu Leu Thr Glu Ala Leu Lys Gln

AAG ATG GGC TTC GAC GGT TTC GTG GTG TCC GAC TGG AAC GGC ATC GGC CAG

Lys Met Gly Phe Asp Gly Phe Val Val Ser Asp Trp Asn Gly Ile Gly Gln

GTC ACC ACC GAG AAC AGC AAC GCG ACG CGC AAC TGC AGC AAC AGC GAC TGC

Val Thr Thr Glu Asn Ser Asn Ala Thr Arg Asn Cys Ser Asn Ser Asp Cys

CCC GAG GCC ATC AAC GCT GGC ATC GAC ATG GTG ATG GTG CCG TAC CGG GCC

Pro Glu Ala Ile Asn Ala Gly Ile Asp Met Val Met Val Pro Tyr Arg Ala

GAC TGG AAG GCC TTC ATC ACC AAC ACA ATT GCA ATT GTC CGC AAA GGC GAG

Asp Trp Lys Ala Phe Ile Thr Asn Thr Ile Ala Ile Val Arg Lys Gly Glu

ATC GCG CAG GAG CGC ATC GAC AAC GCG GTG CGG CGC ATC CTG CGC GTC AAG

Ile Ala Gln Glu Arg Ile Asp Asn Ala Val Arg Arg Ile Leu Arg Val Lys

TTG CGC GCC GGT CTG TTC GAC AAG CCC ACA CCC TCC GCC CGT CTG GCC TCG

Leu Arg Ala Gly Leu Phe Asp Lys Pro Thr Pro Ser Ala Arg Leu Ala Ser

CGC GAG GTC GGC AGC GCC GAA CAC CGG GCG CTC GCG CGT GAA GCG GTG CGC

Arg Glu Val Gly Ser Ala Glu His Arg Ala Leu Ala Arg Glu Ala Val Arg

AAG TCG TTG GTG CTG TTG AAG AAC AAC GGC CGG GTG CTG CCG CTG GCA CGC

Lys Ser Leu Val Leu Leu Lys Asn Asn Gly Arg Val Leu Pro Leu Ala Arg

AAT GCC AAG GTC CTG GTG GCC GGC AAG AGC GCC AAC AGC CTC GAG AAC CAG

Asn Ala Lys Val Leu Val Ala Gly Lys Ser Ala Asn Ser Leu Glu Asn Gln

ACC GGC GGC TGG TCG CTC AGC TGG CAA GGC ACC GGC AAC GCC AAC GCC GAT

Thr Gly Gly Trp Ser Leu Ser Trp Gln Gly Thr Gly Asn Ala Asn Ala Asp

TTC GGC GGC GGC ACG ACC GTG TGG CAG GCG ATC CAG AAG ATC GCC CCG AAT

Phe Gly Gly Gly Thr Thr Val Trp Gln Ala Ile Gln Lys Ile Ala Pro Asn

GCC GAA CTC GAC ACC AGC GCC GAC GGC GCC AAG GGC AGC GAT GCC TAC GAC

Ala Glu Leu Asp Thr Ser Ala Asp Gly Ala Lys Gly Ser Asp Ala Tyr Asp

GCC GCG ATC GTC GTG ATC GGT GAA ACA CCG TAC GCC GAA GGT GTC GGA GAC

Ala Ala Ile Val Val Ile Gly Glu Thr Pro Tyr Ala Glu Gly Val Gly Asp

ATC GGC CGC AGC AAG ACG CTG GAA CTC ACC AAG CTG CGT CCA GAA GAC CTC

Ile Gly Arg Ser Lys Thr Leu Glu Leu Thr Lys Leu Arg Pro Glu Asp Leu

GCC GTG ATC GAA GGC CTG CGC GCC AAG GGC GTG AAG AAA ATC GTC ACG CTG

Ala Val Ile Glu Gly Leu Arg Ala Lys Gly Val Lys Lys Ile Val Thr Leu

CTG GTC TCC GGC CGC CCG CTC TAC GTC AAC AAG GAG CTG AAC CGC TCG GAC

Leu Val Ser Gly Arg Pro Leu Tyr Val Asn Lys Glu Leu Asn Arg Ser Asp

GCC TTC GTG GCG GCG TGG CTG CCC GGC ACC GAA GGC GAC GGC GTC GCC GAC

Ala Phe Val Ala Ala Trp Leu Pro Gly Thr Glu Gly Asp Gly Val Ala Asp

GTG CTG TTC CGT GCG GCC GAC GGC AGC GTC GCG CAT GGC TTC AGC GGC AAG

Val Leu Phe Arg Ala Ala Asp Gly Ser Val Ala His Gly Phe Ser Gly Lys

CTG TCG TTC TCG TGG CCG AAG TCG GCC TGC CAG ACG CCG CTC AAC CGT GGC

Leu Ser Phe Ser Trp Pro Lys Ser Ala Cys Gln Thr Pro Leu Asn Arg Gly

GAC GCC ACC TAC GAC CCG CTC TAC GCT TAT GGC TAC GGC CTT CAA TAC GGC

Asp Ala Thr Tyr Asp Pro Leu Tyr Ala Tyr Gly Tyr Gly Leu Gln Tyr Gly

GAG GAG ACC GAT CAG AGC GCG TAC GAC GAA AGC AGT GCC ACG GTC GGC TGC

Glu Glu Thr Asp Gln Ser Ala Tyr Asp Glu Ser Ser Ala Thr Val Gly Cys

GGC ATC CAG GAC GGC GGC GGC ACC ACG GCC GAG CCG CTG GCG GTG TTC GAA

Gly Ile Gln Asp Gly Gly Gly Thr Thr Ala Glu Pro Leu Ala Val Phe Glu

GGC GGA GCC AAC CAG GGC AAC TGG AAG CTG CGC ATC GGC GCC GAG TCG AGC

Gly Gly Ala Asn Gln Gly Asn Trp Lys Leu Arg Ile Gly Ala Glu Ser Ser

TGG AGC AAC GAT GTG ACG CTG GCC AGC AGC GCG GTG ACG TCG ACG CCG TCC

Trp Ser Asn Asp Val Thr Leu Ala Ser Ser Ala Val Thr Ser Thr Pro Ser

AAC GAA CTG CAG GCC GTG CCG GTG GAC GAC AAG GCC GGG CGG CAA TGG GCG

Asn Glu Leu Gln Ala Val Pro Val Asp Asp Lys Ala Gly Arg Gln Trp Ala

GCG GTG AAG GCG ACC TGG AAC GAC AAG CCC GGC CAG CTC TAC ATG CAA AGC

Ala Val Lys Ala Thr Trp Asn Asp Lys Pro Gly Gln Leu Tyr Met Gln Ser

GCC AAC CCC GGC GAC CTG GTG GAC CTG ATG GCC TAT CAG AAC TCC GGT GGC

Ala Asn Pro Gly Asp Leu Val Asp Leu Met Ala Tyr Gln Asn Ser Gly Gly

GCG CTG GTG TTC GAC CTG CGT GTC GTC AGT GCG CCG ACC GAC CCG GTC AAG

Ala Leu Val Phe Asp Leu Arg Val Val Ser Ala Pro Thr Asp Pro Val Lys

CTG CGC GTC GAT TGC GGC TGG CCC TGT CTG GGC GAG ATC GAC GTC ACC AGC

Leu Arg Val Asp Cys Gly Trp Pro Cys Leu Gly Glu Ile Asp Val Thr Ser

GCC GTC AAG GCC CAG CCG GTC AAC GCC TGG AAG GAA GTG GCG GTG TCG CTG

Ala Val Lys Ala Gln Pro Val Asn Ala Trp Lys Glu Val Ala Val Ser Leu

CAG TGT TTC GCC GAC GCC GGC ACC GAC CTG GCC ATC GTC AAC ACG CCC TTC

Gln Cys Phe Ala Asp Ala Gly Thr Asp Leu Ala Ile Val Asn Thr Pro Phe

CTG ATG TAC ACG TCT GGC CGC TTC GAA GCT GCC GTC GCG AAC ATC CGT TGG

Leu Met Tyr Thr Ser Gly Arg Phe Glu Ala Ala Val Ala Asn Ile Arg Trp

GAG CCC AAG CGC ACG CCC AAC GTG GGG TGC AAC GGC GCA CCG ATC GCC GCC

Glu Pro Lys Arg Thr Pro Asn Val Gly Cys Asn Gly Ala Pro Ile Ala Ala

GCG CCT TGA 2711
Ala Pro END

Pyrococcus furiosus (Clone # 7EG1) Glycosidase

1

ATG AGC AAG AAA AAG TTC GTC ATC GTA TCT ATC TTA ACA ATC CTT TTA GTA

Met Ser Lys Lys Lys Phe Val Ile Val Ser Ile Leu Thr Ile Leu Leu Val

CAG GCA ATA TAT TTT GTA GAA AAG TAT CAT ACC TCT GAG GAC AAG TCA ACT

Gln Ala Ile Tyr Phe Val Glu Lys Tyr His Thr Ser Glu Asp Lys Ser Thr

TCA AAT ACC TCA TCT ACA CCA CCC CAA ACA ACA CTT TCC ACT ACC AAG GTT

Ser Asn Thr Ser Ser Thr Pro Pro Gln Thr Thr Leu Ser Thr Thr Lys Val

CTC AAG ATT AGA TAC CCT GAT GAC GGT GAG TGG CCA GGA GCT CCT ATT GAT

Leu Lys Ile Arg Tyr Pro Asp Asp Gly Glu Trp Pro Gly Ala Pro Ile Asp

AAG GAT GGT GAT GGG AAC CCA GAA TTC TAC ATT GAA ATA AAC CTA TGG AAC

Lys Asp Gly Asp Gly Asn Pro Glu Phe Tyr Ile Glu Ile Asn Leu Trp Asn

ATT CTT AAT GCT ACT GGA TTT GCT GAG ATG ACG TAC AAT TTA ACC AGC GGC

Ile Leu Asn Ala Thr Gly Phe Ala Glu Met Thr Tyr Asn Leu Thr Ser Gly

GTC CTT CAC TAC GTC CAA CAA CTT GAC AAC ATT GTC TTG AGG GAT AGA AGT

Val Leu His Tyr Val Gln Gln Leu Asp Asn Ile Val Leu Arg Asp Arg Ser

AAT TGG GTG CAT GGA TAC CCC GAA ATA TTC TAT GGA AAC AAG CCA TGG AAT

Asn Trp Val His Gly Tyr Pro Glu Ile Phe Tyr Gly Asn Lys Pro Trp Asn

GCA AAC TAC GCA ACT GAT GGC CCA ATA CCA TTA CCC AGT AAA GTT TCA AAC

Ala Asn Tyr Ala Thr Asp Gly Pro Ile Pro Leu Pro Ser Lys Val Ser Asn

CTA ACA GAC TTC TAT CTA ACA ATC TCC TAT AAA CTT GAG CCC AAG AAC GGC

Leu Thr Asp Phe Tyr Leu Thr Ile Ser Tyr Lys Leu Glu Pro Lys Asn Gly

CTG CCA ATT AAC TTC GCA ATA GAA TCC TGG TTA ACG AGA GAA GCT TGG AGA

Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp Leu Thr Arg Glu Ala Trp Arg

ACA ACA GGA ATT AAC AGC GAT GAG CAA GAA GTA ATG ATA TGG ATT TAC TAT

Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val Met Ile Trp Ile Tyr Tyr

GAC GGA TTA CAA CCG GCT GGC TCC AAA GTT AAG GAG ATT GTA GTC CCA ATA

Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu Ile Val Val Pro Ile

ATA GTT AAC GGA ACA CCA GTA AAT GCT ACA TTT GAA GTA TGG AAG GCA AAC

Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val Trp Lys Ala Asn

ATT GGT TGG GAG TAT GTT GCA TTT AGA ATA AAG ACC CCA ATC AAA GAG GGA

Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr Pro Ile Lys Glu Gly

ACA GTG ACA ATT CCA TAC GGA GCA TTT ATA AGT GTT GCA GCC AAC ATT TCA

Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser Val Ala Ala Asn Ile Ser

AGC TTA CCA AAT TAC ACA GAA CTT TAC TTA GAG GAC GTG GAG ATT GGA ACT

Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu Glu Asp Val Glu Ile Gly Thr

GAG TTT GGA ACG CCA AGC ACT ACC TCC GCC CAC CTA GAG TGG TGG ATC ACA

Glu Phe Gly Thr Pro Ser Thr Thr Ser Ala His Leu Glu Trp Trp Ile Thr

AAC ATA ACA CTA ACT CCT CTA GAT AGA CCT CTT ATT TCC TAA 960

Asn Ile Thr Leu Thr Pro Leu Asp Arg Pro Leu Ile Ser End

Vibrio harveyi (Clone # 91GP2) Glycosidase

1
ATG AGA GGT AAC ACG ATG AAG CAA AAA GCG CTA TAT CGA GCA GTA GCA ATG
Met Arg Gly Asn Thr Met Lys Gln Lys Ala Leu Tyr Arg Ala Val Ala Met

GGT TTG AGT GGT CTT GCG AAC GTC GCA TCC GCT AAT GAG ATG GTA AAT CCT
Gly Leu Ser Gly Leu Ala Asn Val Ala Ser Ala Asn Glu Met Val Asn Pro

GAT GGT GGT GTC GTA GTG GGT TAC TGG CAT AAC TGG TGC GAT GGC GCT GGT
Asp Gly Gly Val Val Val Gly Tyr Trp His Asn Trp Cys Asp Gly Ala Gly

TAC AAG GGA GGT AAT GCA CCG TGT GTA ACA TTG GAT GAA GTT GAT CCT ATG
Tyr Lys Gly Gly Asn Ala Pro Cys Val Thr Leu Asp Glu Val Asp Pro Met

TAC AAT GTG GTT AAC GTC TCC TTT ATG AAG GTA TTC AAT ACC AGT GAA GGT
Tyr Asn Val Val Asn Val Ser Phe Met Lys Val Phe Asn Thr Ser Glu Gly

CGT ATT CCA ACC TTT AAG CTC GAT CCA AAT ATC GGC CTT TCA GAA CAA CAA
Arg Ile Pro Thr Phe Lys Leu Asp Pro Asn Ile Gly Leu Ser Glu Gln Gln

TTT TTT GAC CAA ATT GAA GCT CTA AAC CAA CAA GGA CGT GCC GTT CTC ATC
Phe Phe Asp Gln Ile Glu Ala Leu Asn Gln Gln Gly Arg Ala Val Leu Ile

GCT CTT GGT GGC GCA GAT GCT CAC GTT GAA CTT AGA ACT GGT GAC GAA CAA

Ala Leu Gly Gly Ala Asp Ala His Val Glu Leu Arg Thr Gly Asp Glu Gln

GCG TTC GCA CAA GAG ATT ATT CGT TTA ACG GAT AAG TTC GGT TTT GAT GGT

Ala Phe Ala Gln Glu Ile Ile Arg Leu Thr Asp Lys Phe Gly Phe Asp Gly

CTA GAT ATC GAT TTA GAG CAG TCA GCA GTA ACG GCA GAG AAC AAC CAA ACC

Leu Asp Ile Asp Leu Glu Gln Ser Ala Val Thr Ala Glu Asn Asn Gln Thr

GTA ATT CCA GCT GCA CTT CGC CTT GTA AAA GAG CAT TAT CAA CAA CAA GGT

Val Ile Pro Ala Ala Leu Arg Leu Val Lys Glu His Tyr Gln Gln Gln Gly

AAG AAC TTC CTA ATT ACG ATG GCG CCT GAA TTC CCT TAT CTA ACA GAA GGT

Lys Asn Phe Leu Ile Thr Met Ala Pro Glu Phe Pro Tyr Leu Thr Glu Gly

GGC AAG TAT GTT CCT TAC ATT ACT GGT TTA GAA GGG TAC TAC GAT TGG ATC

Gly Lys Tyr Val Pro Tyr Ile Thr Gly Leu Glu Gly Tyr Tyr Asp Trp Ile

AAC CCT CAG TTT TAC AAT CAA GGT GGT GAC GGT ATT TGG GTT GAT GGC GTG

Asn Pro Gln Phe Tyr Asn Gln Gly Gly Asp Gly Ile Trp Val Asp Gly Val

GGT TGG ATA GCG CAA AAC AAT GAT GAG TTA AAA CAA GAG TTT ATT TAC TAC

Gly Trp Ile Ala Gln Asn Asn Asp Glu Leu Lys Gln Glu Phe Ile Tyr Tyr

ATT TCG GAC GCT CTA TCG AAC GGT ACA CGC GGT TTC CAC AAA ATC CCG CAT

Ile Ser Asp Ala Leu Ser Asn Gly Thr Arg Gly Phe His Lys Ile Pro His

GAC AAA CTG GTG TTT GGT ATC CCA TCT AAC ATT GAT GCT GCT GCA ACG GGC

Asp Lys Leu Val Phe Gly Ile Pro Ser Asn Ile Asp Ala Ala Ala Thr Gly

TTT GTT CAA AAC CCT CAA GAC CTT TAC GAC GCG TTT GAT CAA CTT AAA GCG

Phe Val Gln Asn Pro Gln Asp Leu Tyr Asp Ala Phe Asp Gln Leu Lys Ala

CAA GGG CAG GCA CTT CGT GGC GTA ATG ACA TGG TCG GTG AAC TGG GAT ATG

Gln Gly Gln Ala Leu Arg Gly Val Met Thr Trp Ser Val Asn Trp Asp Met

GGC ACC GAT AAA AAT GGC CAA GCG TAC GGT GAA AAA TTC GTG AAG GAT TAC

Gly Thr Asp Lys Asn Gly Gln Ala Tyr Gly Glu Lys Phe Val Lys Asp Tyr

GGT CCG TTT ATC CAC GGG CAG ACT CCA CCA CCA AGT GAA GGT GAA CCA GTT

Gly Pro Phe Ile His Gly Gln Thr Pro Pro Pro Ser Glu Gly Glu Pro Val

TTT AGT GGC CTC AAC GAT GTT CGT GTG CAT CAC GGT AGT TCA TTT GAC CCG

Phe Ser Gly Leu Asn Asp Val Arg Val His His Gly Ser Ser Phe Asp Pro

TAT GCA GGT GTT ACT GCG TCT GAT AAA GAA GAT GGA GAC CTA ACC AAC AGC

Tyr Ala Gly Val Thr Ala Ser Asp Lys Glu Asp Gly Asp Leu Thr Asn Ser

ATC ACT GTC GAA GGT TCA GTT GAT GTG AAC ACG GTA GGC ACA TAT GTT TTG
Ile Thr Val Glu Gly Ser Val Asp Val Asn Thr Val Gly Thr Tyr Val Leu

GTT TAC AGT GTA AAA GAC AGC GAC AAC AAT GAA ACC AAG CAA AGT AGA ACG
Val Tyr Ser Val Lys Asp Ser Asp Asn Asn Glu Thr Lys Gln Ser Arg Thr

GTT GTT GTT TAC AGC CTA GTG CCT GAG TTT GAA GGT GTC GCA GAT ACG ACC
Val Val Val Tyr Ser Leu Val Pro Glu Phe Glu Gly Val Ala Asp Thr Thr

ATC CAG CTT GGT GAC GCT TTT GAC CCA ATG GCA GGC GTA AAA GCG ACG GAT
Ile Gln Leu Gly Asp Ala Phe Asp Pro Met Ala Gly Val Lys Ala Thr Asp

GCA GAA GAC GGT GAT TTG ACT GAT CGG TAT CTA CGC CGC CTA AGG TCA CTT
Ala Glu Asp Gly Asp Leu Thr Asp Arg Tyr Leu Arg Arg Leu Arg Ser Leu

CTG CGG TGC GAT AGC CTT CTG TGC CAT TTG GTG CAA CCG CCC AGT TTT CCA
Leu Arg Cys Asp Ser Leu Leu Cys His Leu Val Gln Pro Pro Ser Phe Pro

GAC GCT CAA CGA TGG TTG CCA TCT CTT TCT GGT TGA 1514
Asp Ala Gln Arg Trp Leu Pro Ser Leu Ser Gly END

DECLARATION FOR PATENT APPLICATION

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship is as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled ENDOGLUCANASES, the specification of which

____ is attached hereto.

__X__ was filed on April 18, 2001 as U.S. Application Serial No. 09/914,543 (the "Application") (Docket No. DIVER1150-5

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the Application as defined in Title 37, Code of Federal Regulations ("C.F.R."), § 1.56.

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

_____ _____

(Application Serial No.) (Filing Date)

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of the application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the

Application as defined in Title 37, C.F.R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of the Application:

(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

COUNTRY	APPLICATION NO.	FILING DATE	PRIORITY CLAIMED	
US	PCT/US97/08793	May 22, 1997	■ Yes	□ No

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

1-CO

Full name of first inventor: David E. Lam

Inventor's signature: [signature]

Date: 1/14/2002

Residence: Carlsbad, California CA

Citizenship: USA

Post Office Address: 3261 Avenida Anacapa
Carlsbad, California 92009

[signature]

2-00

Full name of second inventor: Eric J. Mathur

Inventor's signature: [signature]

Date: 1/16/01

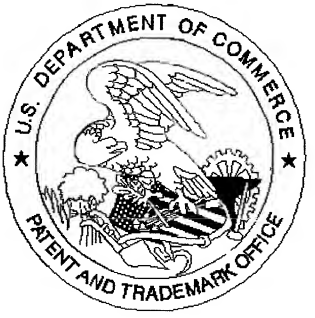
Residence: Carlsbad, California CA

Citizenship: USA

Post Office Address: 2654 Galicia Way
Carlsbad, California 92009

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☑ Page(s)_____ of Drawing _____ were not present for scanning.
(Document title)

☐ Page(s)_____ of _____ were not present for scanning.
(Document title)

☐ ***Scanned copy is best available.***